ROGER PENROSE

La mente nueva del emperador

En torno a la cibernética, la mente y las leyes de la física



CONSEJO NACIONAL DE CIENCIA Y TECNOLOGIA

FONDO DE CULTURA ECONÓMICA

MÉXICO

Traducción:

JOSÉ JAVIER GARCÍA SANZ

Primera edición en inglés, 1989

Primera edición en español, 1996

Primera reimpresión, 1996

Título original:

The Emperor's New Mind —Concerning Computers, Minds, and The Laws of Physics

ISBN 0-19-851973-7

Carretera Picacho-Ajusco 227, 14200 México, D.F.

ISBN 968-16-4361-5

Impreso en México

Dedico este libro a la memoria de mi querida madre,
quien no vivió para verlo.

NOTA PARA EL LECTOR:

sobre la lectura de las ecuaciones matemáticas

En diferentes partes de este libro he recurrido al uso de ecuaciones matemáticas, desoyendo impertérrito las frecuentes advertencias de que cada una de estas fórmulas reduciría a la mitad el número de lectores. Si usted es una persona que se siente intimidada ante una fórmula (como la mayoría de la gente), entonces le recomiendo un método que yo mismo uso cuando se presenta una de estas fórmulas fastidiosas. El método consiste, más o menos, en pasarla por alto y saltar a la siguiente línea de texto. Bien, no exactamente; es conveniente echar una rápida ojeada a la pobre fórmula, sin tratar de comprenderla del todo, y luego seguir adelante. Algún tiempo después, y armados con nueva confianza, podemos volver a la fórmula olvidada y tratar de captar alguna de sus características más sobresalientes. El propio texto puede servir de ayuda para saber qué es lo importante y qué puede ser pasado por alto sin problemas. Si no lo consigue, entonces prescinda de la fórmula, por completo y sin remordimientos.

AGRADECIMIENTOS

Muchas personas me han ayudado, de una u otra forma, a escribir este libro, y debo darles las gracias. En particular, a los defensores de la IA fuerte (especialmente a los que intervinieron en un programa de la BBC TV que tuve ocasión de presenciar), quienes al expresar opiniones tan radicales me incitaron, hace ya varios años, a embarcarme en este proyecto. (Pese a todo, me temo que si entonces hubiera sabido el esfuerzo que su escritura me iba a exigir, no lo hubiera empezado.)

También quiero agradecer a Toby Bailey, David Deutsch (quien también fue de gran ayuda en la comprobación de las especificaciones de mi máquina de Turing), Stuart Hampshire, Jim Hartle, Lane Hughston, Angus McIntyre, Mary Jane Mowat, Tristan Needham, Ted Newman, Eric Penrose, Toby Penrose, Wolfgang Rindler, Engelbert Schücking y Dennis Sciama, quienes revisaron versiones de pequeñas partes del manuscrito y me hicieron muchas sugerencias valiosas para mejorarlo.

Merece un reconocimiento especial la ayuda de Christopher Penrose, con la información detallada respecto al conjunto de Mandelbrot, así como la de Jonathan Penrose, por su valiosa información sobre la computadora que juega ajedrez. Muchas gracias también a Colin Blakemore, Erich Harth y David Hubel por leer y revisar el capítulo IX, que concierne a un tema en el que sinceramente no soy un experto aunque, como sucede con toda la gente que acabo de mencionar, ellos no son en absoluto responsables de los errores que puedan haber quedado. Agradezco a la NSF su ayuda mediante los contratos DMS 84-05644, DMS 86-06488 y PHY 86-12424. Asimismo he contraído una gran deuda con Martin Gardner por su extrema generosidad al escribir el prefacio de este trabajo, y también por sus comentarios. De forma especial, quiero agradecer a mi querida Vanessa por sus atentas y detalladas críticas en varios capítulos, por su invaluable asistencia en las referencias y, lo más importante, por soportarme cuando era insoportable, por su cariño y apoyo donde y cuando era vital.

PROCEDENCIA DE LAS ILUSTRACIONES

Los editores agradecen el permiso para reproducir las ilustraciones que se citan:

Las figs. IV.6 y IV.9 proceden de D. A. Klarner (ed.) *The mathematical Gardner* (Wadsworth International, 1981).

La fig. IV.7 procede de B. Grünbaum y G. C. Shephard, *Tilings and patterns* (W. H. Freeman, 1987). Copyright © 1987 por W. H. Freeman and Company. Utilizada con su permiso.

La fig. IV.10 procede de K. Chandrasekharan, *Hermann Weyl 1885-1985* (Springer, 1986).

Las figs, IV.11 y X.3 proceden de Pentaplexity: a class of non-periodic tilings of the plane. *The Mathematical Intelligencer*, *2*, 32-7 (Springer, 1979).

La fig. IV.12 procede de H. S. M. Coxeter, M. Emmer, R. Penrose y M. L. Teuber (eds.) y M. C. Escher: *Art and science* (North Holland, 1986).

La fig. V.2 © 1989 M. C. Escher Heirs/Cordon Art—Baarn—Holland.

La fig. X.4 procede de *Journal of Materials Research*, *2*, 1-4 (Materials Research Society, 1987).

Todas las demás ilustraciones son del autor.

PREFACIO

***por* MARTIN GARDNER**

PARA muchos matemáticos y físicos célebres resulta difícil, si no imposible, escribir un libro que pueda ser entendido por los profanos. Hasta hoy se podría haber pensado que Roger Penrose, uno de los físico-matemáticos más eruditos y creativos del mundo, pertenecía a esta clase. Aunque quienes habíamos leído sus artículos y conferencias de divulgación teníamos otra opinión. Aun así, fue una deliciosa sorpresa descubrir que Penrose había robado tiempo a sus ocupaciones para producir un libro maravilloso destinado al profano. Creo que este libro pronto será clásico.

Aunque los capítulos del libro de Penrose recorren la teoría de la relatividad, la mecánica cuántica y la cosmología, su interés principal radica en lo que los filósofos llaman el "problema mente-cuerpo". Durante décadas los defensores de la "IA (Inteligencia Artificial) fuerte" han intentado convencernos de que sólo es cuestión de uno o dos siglos (algunos hablan incluso de cincuenta años), para que las computadoras electrónicas hagan todo lo que la mente humana puede hacer. Estimulados por lecturas juveniles de ciencia-ficción y convencidos de que nuestras mentes son simplemente "computadoras hechas de carne" (como Marvin Minsky dijo en cierta ocasión), dan por supuesto que el placer y el dolor, el gusto por la belleza, el sentido del humor, la conciencia y el libre albedrío son cualidades que emergerán de modo natural cuando el comportamiento algorítmico de los robots electrónicos llegue a ser suficientemente complejo.

Algunos filósofos de la ciencia (en particular John Searle, cuyo famoso experimento mental de la habitación china discute Penrose en detalle) están en abierto desacuerdo. Para ellos una computadora no es esencialmente diferente de las calculadoras mecánicas que funcionan con ruedas, palancas o cualquier otro mecanismo que transmita señales. (Se puede construir una computadora a base de ruedas que giran o agua que se mueva por tuberías.) Puesto que la electricidad viaja por los cables conductores mucho más rápido que otras formas de energía (excepto la luz), también puede jugar con los símbolos más rápidamente que las calculadoras mecánicas, y realizar así tareas de enorme complejidad. Pero ¿"comprende" una computadora electrónica lo que está haciendo en una medida superior a la "comprensión" de la que es capaz un ábaco? Las computadoras juegan ahora al ajedrez como un gran maestro. ¿"Comprenden" su juego mejor de lo que lo hace la máquina de jugar a *tres en raya* que en cierta ocasión construyeron unos desguazadores de computadoras con piezas de chatarra?

Este libro es el ataque más poderoso que se haya escrito contra la IA fuerte. Durante los últimos siglos se han levantado objeciones contra el alegato reduccionista de que la mente es una máquina que funciona según las conocidas leyes de la física, pero la ofensiva de Penrose es más convincente, puesto que hace uso de información de la que no disponían los escritores anteriores. En el libro, Penrose se revela como algo más que un físico-matemático: es también un filósofo de primera línea, que no teme abordar problemas que sus contemporáneos despachan considerándolos sin sentido.

Penrose tiene también el valor de sostener, frente al creciente rechazo de un pequeño grupo de físicos, un vigoroso realismo. No sólo el Universo "está ahí", sino que la verdad matemática tiene también sus propias y misteriosas independencia e intemporalidad. Como Newton y

Einstein, Penrose tiene un profundo sentido de humildad y respeto tanto hacia el mundo físico como hacia el ámbito platónico de la matemática pura. Al famoso especialista en teoría de números Paul Erdós le gusta hablar del "libro de Dios" en el que están registradas las demostraciones más notables. A los matemáticos se les permite de cuando en cuando echar una ojeada a alguna página. Penrose cree que cuando un físico o un matemático experimenta una repentina ¡*eureka*!, no se trata simplemente de algo "producido por un cálculo complicado": es que la mente, por un momento, entra en contacto con la verdad objetiva. ¿No sería posible, se pregunta, que el mundo de Platón y el mundo físico (que los físicos han diluido ahora en las matemáticas) fueran realmente uno y el mismo?

Muchas páginas del libro están dedicadas a la famosa estructura de tipo fractal conocida como conjunto de Mandelbrot, por ser Benoit Mandelbrot quien la descubrió. Aunque es autosimilar en sentido estadístico, a medida que sus partes son ampliadas, su estructura con infinitas circunvoluciones cambia de manera impredecible. Penrose encuentra incomprensible (igual que yo) que nadie pueda suponer que esta exótica estructura no "esté ahí" igual que lo está el monte Everest, y pueda ser explorada de la misma forma que se explora una selva.

Penrose forma parte del cada vez mayor grupo de físicos que piensan que Einstein no era ni tan obstinado ni tan confuso cuando afirmaba que una "voz interior" le decía que la mecánica cuántica estaba incompleta. Para apoyar esta afirmación, Penrose lleva al lector por un fascinante recorrido a través de temas como los números complejos, las máquinas de Turing, la teoría de la complejidad, las desconcertantes paradojas de la mecánica cuántica, los sistemas formales, la indecidibilidad de Gödel, los espacios fase, los espacios de Hilbert, los agujeros negros, los agujeros blancos, la radiación de Hawking, la entropía o la estructura del cerebro, y tantea otros temas que están en el centro de las especulaciones actuales. ¿Tienen los perros y los gatos "conciencia" de sí mismos? ¿Es posible, en teoría, para una máquina que transmite materia, transferir a una persona de un lugar a otro de la misma manera en que eran transmitidos o recibidos los astronautas de una serie de televisión? ¿Cómo ayudó a la supervivencia el que la evolución haya producido la conciencia? ¿Existe un nivel más allá de la mecánica cuántica en el que la dirección del tiempo y la distinción entre izquierda y derecha estén indisolublemente asociados? ¿Son las leyes de la mecánica cuántica —o quizás otras leyes aún más profundas— esenciales para la actuación de la mente?

La respuesta de Penrose a las dos últimas preguntas es afirmativa. Su famosa teoría de los *twistors* —objetos geométricos abstractos que operan en un espacio complejo multidimensional que subyace bajo el espacio-tiempo— es demasiado técnica para ser incluida en este libro. Ellos representan los esfuerzos de Penrose durante dos décadas para sondear una región más profunda que la de los campos y las partículas en la mecánica cuántica. Al clasificar las teorías en cuatro categorías: extraordinarias, útiles, provisionales y erróneas, Penrose coloca modestamente su teoría de los *twistors* en la clase de las provisionales, junto con la de las supercuerdas u otros grandes esquemas unificadores que hoy son fuertemente debatidos.

Penrose es, desde 1973, el catedrático Rouse Ball de Matemáticas en la Universidad de Oxford. El título es apropiado ya que W. W. Rouse Ball no sólo fue un notable matemático sino también un mago aficionado, con un interés tan apasionado por las matemáticas recreativas que escribió una obra clásica en este campo: *Mathematical Recreations and Essays.* Penrose comparte el entusiasmo de Ball por el juego. En su juventud descubrió un "objeto imposible" llamado "tribar". (Un objeto imposible es el dibujo de una figura sólida que no puede existir ya que

incorpora elementos contradictorios.) Él y su padre Lionel, genetista, convirtieron el tribar en la Escalera de Penrose, una estructura que Maurits Escher utilizó en dos famosas litografías: "Ascenso y Descenso" y "Cascada". Un día en que Penrose estaba tumbado en la cama imaginó, en lo que él llamó un "arrebato de locura", un objeto imposible en un espacio tetradimensional. Es algo, decía, que si se le mostrase a una criatura del espacio de cuatro dimensiones le haría exclamar: "¿qué es esto? ¡Dios mío!"

Durante los años sesenta, mientras trabajaba en cosmología con su amigo Stephen Hawking, Penrose hizo el que tal vez sea su descubrimiento más conocido. Si la teoría de la relatividad es válida "hasta el final", en todo agujero negro debe haber una singularidad en la que ya no sean aplicables las leyes de la física. Incluso este resultado ha sido eclipsado últimamente por la construcción que él mismo hizo de dos formas que embaldosan el plano, a la manera de la teselación de Escher, pero que sólo pueden hacerlo en forma no periódica. (Encontrará una discusión de estas sorprendentes formas en mi libro *Penrose Tiles to Trapdoor Ciphers.*) Penrose las inventó, o más bien las descubrió, sin esperar que fueran de utilidad. Para asombro de todos resultó que las formas tridimensionales de sus baldosas pueden subyacer bajo un extraño y nuevo tipo de materia. El estudio de estos "cuasicristales" es hoy en día una de las áreas de investigación más activas dentro de la cristalografía. Es también el ejemplo más espectacular en los tiempos modernos de cómo las matemáticas lúdicas pueden tener aplicaciones no previstas.

Los resultados de Penrose en matemáticas y física —y sólo he mencionado una pequeña parte— surgen de una permanente admiración por el misterio y por la belleza del ser. Su voz interior le dice que la mente humana es algo más que una simple colección de minúsculos cables e interruptores. El Adam de sus prólogo y epílogo es en parte el símbolo del despertar de la conciencia en la lenta evolución de la vida sensible. Para mí, Penrose es también el niño sentado en la tercera fila, detrás de las vacas sagradas de la IA, y que se atreve a sugerir que el emperador de la IA fuerte va desnudo. Aunque sus opiniones estén salpicadas de humor, ésta no es materia de risa.

PRÓLOGO

Hay una numerosa concurrencia en el Gran Auditorio para asistir a la presentación de la nueva computadora "Ultronic". El presidente Polho acaba de concluir su discurso de apertura y se alegra de ello: no se siente a gusto en tales ocasiones y no sabe nada de computadoras, salvo que ésta le va a ahorrar mucho tiempo. Los fabricantes le han asegurado que, entre sus muchos cometidos, será capaz de asumir todas las delicadas decisiones de Estado tan fastidiosas para él. Mejor que así sea, considerando la cuantiosa suma que se ha invertido en ello. Se ve ya disfrutando de muchas horas libres para jugar al golf en su magnífico campo privado, una de las pocas áreas verdes extensas que quedan en su pequeño país.

Adam se sentía privilegiado de contarse entre los asistentes a la ceremonia de inauguración. Se sentó en la tercera fila. Dos filas más adelante de él estaba su madre, tecnócrata que había intervenido en el diseño de Ultronic. Casualmente su padre también estaba allí —en el fondo de la sala—, completamente rodeado de guardias de seguridad. En el último minuto el padre de Adam había tratado de hacer estallar la computadora. Él mismo se había encomendado esta misión, autonombrándose "espíritu conductor" de un pequeño grupo de activistas: el Gran Consejo para la Conciencia Psíquica. Por supuesto, él con todos sus explosivos habían sido inmediatamente detectados por los numerosos sensores electrónicos y químicos. Una pequeña parte de su castigo consistiría en ser testigo de la ceremonia de inauguración.

Adam no sentía especial aprecio por sus padres. Quizá no necesitaba tales sentimientos. Durante sus trece años había sido criado casi exclusivamente por computadoras rodeado de todas las comodidades. Podía tener todo lo que quisiera sin más que apretar un botón: comida, bebida, compañía y entretenimiento; también información sobre cualquier cosa que le interesara, siempre ilustrada con coloridas y atractivas ilustraciones. La alta posición de su madre había hecho posible todo esto.

El diseñador en jefe estaba llegando al final de *su* discurso: "...tiene más de 10^{17} unidades lógicas. ¡Más que el número total de neuronas que reúnen todos los cerebros de todas las personas en todo el país! Su inteligencia será inimaginable. Afortunadamente, sin embargo, no necesitarnos imaginarla. Dentro de un instante todos nosotros tendremos el Privilegio de ser testigos de primera mano de su inteligencia: ¡pido a la respetable primera dama de nuestro gran país, la señora Isabella Polho, que conecte el interruptor que activa nuestra fantástica computadora Ultronic!"

La esposa del presidente avanzó. Un poco nerviosa, y con cierta torpeza, cerró el interruptor. Se produjo un gran silencio y un casi imperceptible parpadeo de las luces cuando las 10^{17} unidades lógicas se activaron. Todos esperaban, sin saber muy bien el qué. "Bien, ¿hay alguien en la audiencia que quiera dirigirse a nuestro nuevo Sistema de Cómputo Ultronic para plantearle la primera pregunta?", interrogó el diseñador en jefe. Nadie se atrevía, temerosos de parecer estúpidos ante la multitud, y ante la nueva omnipresencia. Se hizo el silencio. "Sin duda hay alguien", suplicó. Pero todos tenían miedo, aprensivos frente a la nueva y todopoderosa conciencia. Pero Adam no sentía el mismo respeto, por el hecho de haber crecido entre computadoras. Casi sabía lo que se sentiría *ser* una computadora, o por lo menos así lo creía. De todas formas tenía curiosidad. Levantó su mano. "Ah, sí", dijo el diseñador en jefe, "el muchacho de la tercera fila. ¿Tienes alguna pregunta para nuestro —ejem— nuevo amigo?"

I. ¿CABE LA MENTE EN UNA COMPUTADORA?

INTRODUCCIÓN

DURANTE LAS ÚLTIMAS DÉCADAS, la tecnología de las computadoras electrónicas ha hecho enormes progresos. Y estoy seguro de que en las próximas décadas tendrán lugar nuevos progresos en velocidad, capacidad y diseño lógico. Nuestras computadoras actuales nos parecerán tan lentas y primitivas como hoy nos lo parecen las calculadoras mecánicas de antaño. Hay algo casi estremecedor en el ritmo del progreso. Las computadoras ya pueden realizar con mucha más velocidad y precisión tareas que hasta ahora habían estado reservadas exclusivamente al pensamiento humano. Desde hace tiempo estamos acostumbrados a que las máquinas nos superen ampliamente en las tareas *físicas. Esto* no nos causa el menor desasosiego. Antes bien, nos gusta tener aparatos que nos lleven por tierra a grandes velocidades —más de cinco veces la velocidad del más veloz atleta humano— o que puedan cavar hoyos o demoler estructuras que nos estorban con una rapidez que dejaría en ridículo a equipos compuestos por docenas de hombres. Estamos aún más encantados de tener máquinas que nos permitan hacer físicamente cosas que nunca antes habíamos podido hacer, como llevarnos por los cielos y depositarnos al otro lado del océano en cuestión de horas. El que las máquinas obtengan tales logros no hiere nuestro orgullo. Pero el *poder pensar*, eso sí ha sido siempre una prerrogativa humana. Después de todo, ha sido esa capacidad la que, al traducirse en términos físicos, nos ha permitido superar nuestras limitaciones físicas y la que parecería ponernos por encima de otras criaturas. Si las máquinas pudieran llegar a superarnos algún día en esa cualidad en la que nos habíamos creído superiores, ¿no tendríamos entonces que ceder esa superioridad a nuestras propias creaciones?

La pregunta de si se puede afirmar o no que un artefacto mecánico piensa —quizás incluso que experimenta sentimientos, o que posee una mente—, es antigua.[1] Sin embargo, ha recibido un nuevo ímpetu con la llegada de la moderna tecnología de las .computadoras. Es una pregunta que implica profundos temas de filosofía. ¿Qué significa pensar o sentir? ¿Qué es la mente? ¿Existe realmente la mente? Suponiendo que sí existe, ¿en qué medida depende de las estructuras físicas a las que está asociada? ¿Podría existir la mente al margen de tales estructuras? ¿O es simplemente el modo de funcionar de ciertos tipos de estructuras físicas? En cualquier caso, ¿es imprescindible que las estructuras importantes sean de naturaleza biológica (cerebros) o podrían también estar asociadas a componentes electrónicos? ¿Está la mente sujeta a las leyes de la física? ¿Qué *son*, de hecho, las leyes de la física?

Éstas son algunas de las cuestiones que intentaré tratar en este libro. Pedir respuestas definitivas a preguntas tan fundamentales estaría fuera de lugar. Yo no puedo proporcionar tales respuestas; nadie puede, aunque hay quien trata de impresionarnos con sus conjeturas. Mis propias conjeturas jugarán un papel importante en lo que sigue, pero trataré de distinguir claramente tales especulaciones de los hechos científicos brutos, y trataré también de dejar claras las razones en las que se fundamentan mis especulaciones. No obstante, mi principal propósito aquí no es hacer conjeturas, sino plantear algunos temas aparentemente nuevos, concernientes a la relación entre la estructura de las leyes físicas, la naturaleza de las matemáticas y el pensamiento consciente, y presentar un punto de vista que no he visto expresado hasta ahora. Es un punto de vista que no puedo describir adecuadamente en pocas palabras, y ésta es una de las razones por las que he tenido que realizar un libro de este tamaño. Pero en resumen, y quizá de manera algo equívoca,

---

[1] Véase, por ejemplo, Gardner (1958), Gregory (1981) y las referencias que allí figuran

puedo al menos afirmar que mi punto de vista sugiere que es nuestra actual incomprensión de las leyes fundamentales de la física la que nos impide aprehender el concepto de "mente" en términos físicos o lógicos. No quiero decir con esto que las leyes no sean nunca conocidas del todo. Por el contrario, parte del objetivo de esta obra es intentar estimular la investigación en este campo en direcciones que parecen prometedoras y hacer algunas sugerencias bastante concretas, aparentemente nuevas, sobre el lugar que realmente podría ocupar la mente en el desarrollo de la física que conocemos.

Debería dejar claro que mi punto de vista es poco convencional, al menos entre los físicos y, por consiguiente, resulta poco probable que sea adoptado, actualmente, por los científicos de computadoras o por los fisiólogos. La mayoría de los físicos alegará que las leyes fundamentales que operan a escala del cerebro humano son ya perfectamente conocidas. No se negará, por supuesto, que existen aún muchas lagunas en nuestro conocimiento de la física en general. Por ejemplo, no conocemos las leyes básicas que determinan los valores de la masa de las partículas subatómicas ni la intensidad de sus interacciones. No sabemos cómo hacer del todo compatible la teoría cuántica con la teoría de la relatividad especial de Einstein, ni mucho menos cómo construir la teoría de la "gravitación cuántica" que haga compatible la teoría cuántica con su teoría de la relatividad general. Como consecuencia de esto último, no comprendemos la naturaleza del espacio a la escala absurdamente minúscula de 1/100.000.000.000.000.000.000 del tamaño de las partículas elementales conocidas, aunque para dimensiones mayores nuestro conocimiento se presuma adecuado. No sabemos si el Universo como un todo tiene extensión finita o infinita —tanto en el espacio como en el tiempo— aunque pueda parecer que tales incertidumbres no tengan ninguna importancia en la escala humana. No comprendemos la física que actúa en el corazón de los agujeros negros ni en el *big bang*, origen del propio Universo. Pero todas estas cosas parecen no tener nada que ver con lo que imaginamos en la escala "cotidiana" (o incluso una más pequeña) del funcionamiento del cerebro humano. Y ciertamente así es, aunque argumentaré *precisamente* que en este nivel existe —frente (o, mejor dicho, detrás) de nuestras propias narices— otra gran incógnita en nuestra comprensión de la física y que podría ser fundamental para el funcionamiento del pensamiento humano y de la conciencia. Es una incógnita que no ha sido siquiera reconocida por la mayoría de los físicos, como trataré de demostrar. Argumentaré, además, que curiosamente, los agujeros negros y el *big bang* realmente *tienen* una gran relación con estos asuntos.

En seguida intentaré persuadir al lector de la fuerza de la evidencia que sustenta el punto de vista que trato de exponer. Para comprenderlo, tenemos un buen trabajo por delante. Necesitaremos viajar por territorios muy extraños —algunos de importancia aparentemente dudosa— y por campos de esfuerzo muy distintos. Necesitaremos examinar la estructura, fundamentos y enigmas de la teoría cuántica; los rasgos básicos de las teorías de la relatividad especial y general, de los agujeros negros, del *big bang*, y de la segunda ley de la termodinámica, de la teoría de Maxwell de los fenómenos electromagnéticos y de las bases de la mecánica newtoniana. Además tendremos que vérnoslas con algunas cuestiones de filosofía y psicología cuando intentemos comprender la naturaleza y la función de la conciencia. Por supuesto, tendremos que tener una visión general de la neurofisiología del cerebro, además de los modelos de computadora propuestos. Necesitaremos tener alguna noción del *status* de la inteligencia artificial, así como saber qué es una máquina de Turing, y comprender el significado de la computabilidad, del teorema de Gödel y de la teoría de la complejidad. Nos adentraremos también en los fundamentos de la matemática, e incluso deberemos plantearnos la cuestión de la

propia naturaleza de la realidad física. Si, al final de todo ello, los argumentos menos convencionales que trato de exponer no han persuadido al lector, confío al menos que habrá sacado algo de este tortuoso y, espero, fascinante viaje.

LA PRUEBA DE TURING

Imaginemos que un nuevo modelo de computadora ha salido al mercado, posiblemente con una memoria de almacenamiento y un número de unidades lógicas mayor que las que hay en un cerebro humano. Supongamos también que las máquinas han sido cuidadosamente programadas y que se les ha introducido una gran cantidad de datos. Los fabricantes dirían que el artefacto realmente piensa. Quizá también digan que *es* auténticamente inteligente. O pueden ir más lejos y sugerir que este aparato realmente *siente* dolor, felicidad, compasión, orgullo, etc., y que es consciente y realmente *comprende* lo que está haciendo. De hecho, se está afirmando que tiene *conciencia.*

¿Cómo decidir si son ciertas o no las afirmaciones de los fabricantes? Normalmente, cuando compramos una determinada máquina juzgamos su valor de acuerdo con el servicio que nos presta. Si realiza satisfactoriamente las tareas que le encomendamos, entonces quedamos complacidos. Si no, la devolvemos para su reparación o sustitución. De acuerdo con este criterio, para probar la afirmación de los fabricantes de que un aparato semejante tiene realmente las cualidades humanas que se le atribuyen, pediríamos simplemente que se *comporte*, en estos aspectos, como lo haría cualquier persona. Mientras lo hiciera satisfactoriamente no tendríamos ningún motivo de queja y no necesitaríamos devolver la computadora para su reparación o sustitución.

Esto nos proporciona un punto de vista operacional para abordar estas cuestiones. El conductista dirá que la *computadora piensa* siempre y cuando actúe del mismo modo que lo hace una persona cuando está pensando. Adoptaremos, de momento, este punto de vista operacional. Esto no quiere decir, por supuesto, que estemos pidiendo que la computadora se mueva como podría hacerlo una persona mientras está pensando. Menos aún esperaríamos que se asemejara o se hiciera sentir al tacto como un ser humano: estos atributos serían irrelevantes para el propósito de la computadora. Lo que esto quiere decir, no obstante, es que le estamos pidiendo que dé respuestas de tipo humano a cualquier pregunta que le podamos plantear, y que estamos afirmando que realmente piensa (o siente, comprende, etc.) siempre que responda a nuestras preguntas de una manera indistinguible a la de un ser humano.

Este punto de vista fue vigorosamente expuesto en un famoso artículo titulado "Compunting Machinery and Intelligence", por Alan Turing, aparecido en 1950 en la revista filosófica *Mind* (Turing, 1950). (Hablaremos de Turing más adelante.) En este artículo se describió por primera vez la idea ahora conocida como *prueba de Turing*. Ésta pretendía ser una forma de decidir, dentro de lo razonable, si una máquina efectivamente piensa. Supongamos que se afirma realmente que una computadora (como la que nos venden los fabricantes en la descripción anterior) piensa. De acuerdo con la prueba de Turing, la computadora y algún voluntario humano se ocultan de la vista de un interrogador perspicaz. El interrogador tiene que tratar de decidir cuál es cuál entre la computadora y el ser humano, mediante el simple procedimiento de plantear preguntas de prueba a cada uno de ellos. Estas preguntas y, lo que es más importante, las

respuestas que ella* recibe, se transmiten de modo impersonal; por ejemplo, pulsadas en un teclado o mostradas en una pantalla. A la interrogadora no se le permite más información sobre cualquiera de las partes que la que obtiene en esta sesión de preguntas y respuestas. El sujeto humano responde a las preguntas sinceramente y trata de persuadirle de que él es realmente el ser humano pero la computadora está programada para "mentir" y, por consiguiente, tratar de convencer a la interrogadora de que es ella, y no el otro, el ser humano. Si en el curso de una serie de pruebas semejantes la interrogadora es incapaz de identificar de una forma definitiva al sujeto humano real, se considera que la computadora (su programa, el programador, o el diseñador, etc.) ha superado la prueba.

Ahora bien, si alguien cree que esta prueba es bastante injusta con la computadora, piense por un momento que se invirtieran los papeles de forma que se le pidiese al ser humano que se hiciera pasar por una computadora y a ésta que respondiese sinceramente. Sería demasiado fácil para la interrogadora descubrir cuál es cuál. Todo lo que necesitaría hacer es pedir al sujeto que realizara alguna operación aritmética complicada. Una buena computadora sería *capaz* de responder al instante con precisión, mientras que el ser humano se quedaría mudo. (Habría que tener cierto cuidado con esto, no obstante. Hay humanos "calculadores prodigio" que pueden realizar notables proezas de aritmética mental con precisión infalible y sin esfuerzo aparente. Por ejemplo, Johann Martin Zacharias Dase,[2] hijo de un granjero analfabeto, que vivió en Alemania entre 1824 y 1861, era *capaz* de multiplicar mentalmente dos números de ocho cifras en menos de un minuto, o dos números de veinte cifras en unos seis minutos. Sería fácil confundir tales proezas con los cálculos de una computadora. En tiempos más recientes fueron igualmente impresionantes los logros de Alexander Aitken, que *fue* catedrático de matemáticas en la Universidad de Edimburgo en los años cincuenta, y de algunos otros. La tarea aritmética que escogiera la interrogadora para hacer la prueba tendría que ser mucho más compleja que ésta; por ejemplo, multiplicar dos números de treinta cifras en dos segundos, lo que está claramente dentro de las capacidades de una buena computadora moderna.)

Por consiguiente, parte del trabajo de los programadores de la computadora consiste en hacer que parezca en algunas cosas "más estúpida" de lo que realmente es. Así, si la interrogadora planteara a la computadora una tarea aritmética complicada, como las que hemos considerado más arriba, ésta debería simular que *no* es capaz de responderla o de lo contrario sería descubierta inmediatamente. No creo, sin embargo, que hacer a una computadora "más estúpida" fuera un problema particularmente serio para los programadores. Su dificultad principal estaría en hacer que respondiera a algunos de los tipos de preguntas más simples, de "sentido común", preguntas con las que el sujeto humano no tendría ninguna dificultad.

No obstante, hay una dificultad al dar ejemplos concretos de tales preguntas. Dada cualquier pregunta, sería cosa fácil, a continuación, pensar una manera de hacer que la computadora respondiera a esa pregunta *concreta* como lo haría una persona. Pero cualquier falta de comprensión real por parte de la computadora quedaría probablemente en evidencia en un interrogatorio *continuo*, y especialmente con preguntas originales y que requieran una

* Al escribir un trabajo como éste se presenta un problema inevitable cuando hay que decidir si usar los pronombres "él" o "ella" donde no se tenga intención de referirse al género. Por lo mismo, cuando haga referencia a alguna persona abstracta usaré en lo sucesivo él para indicar simplemente la frase "ella o él", lo que considero una práctica común. Sin embargo, espero que se me perdone un evidente rasgo de sexismo al expresar aquí una preferencia por un interrogador femenino. Mi idea es que ella podría ser más sensible que su contraparte masculina en cuanto a reconocer cualidades humanas reales.

[2] Véase, por ejemplo, Resnikoff y Wells (1984), pp. 181-184. Para un informe clásico sobre los calculadores prodigio en general, véase Rouse Ball (1892); también Smith (1983).

comprensión real. La habilidad de la interrogadora radicaría, en parte, en imaginar estas preguntas originales, y en parte, en hacerlas seguir de otras, de naturaleza exploratoria, diseñadas para descubrir si ha habido o no una "comprensión" real. Podría también plantear de vez en cuando alguna pregunta completamente sin sentido para ver si la computadora puede detectar la diferencia, o bien podría añadir una o dos que superficialmente pareciesen absurdas, pero que en realidad tuviesen cierto sentido. Por ejemplo, podría decir: "Esta mañana oí que un rinoceronte iba volando por el Mississippi en un globo rosa. ¿Qué piensas de eso?" (Casi podemos imaginar las gotas de sudor frío corriendo por la frente de la computadora, por usar una metáfora no muy apropiada). Podría responder cautelosamente: "Me suena bastante ridículo". Hasta aquí todo va bien. Continúa la interrogadora: "¿De veras? Mi tío lo hizo una vez de ida y vuelta, sólo que era beige con rayas. ¿Qué hay de ridículo en eso?" Es fácil imaginar que, si no tuviera una correcta "comprensión", una computadora caería pronto en la trampa y se descubriría. Podría incluso equivocarse y decir: "Los rinocerontes no pueden volar" (si sus bancos de memoria vinieran en su ayuda con el hecho de que no tienen alas), en respuesta a la primera pregunta, o "los rinocerontes no tienen rayas", en respuesta a la segunda. La vez siguiente la interrogadora podría hacer la pregunta más absurda, y cambiarla por "*bajo* el Mississippi", o "en el interior del globo rosa", o "con un camisón rosa", para ver si la computadora tenía juicio para darse cuenta de la diferencia esencial.

Dejemos a un lado, de momento, la cuestión de si puede —o cuándo podría hacerse —construir una computadora que supere realmente la prueba de Turing. En lugar de ello supongamos, sólo para nuestra argumentación, que ya se han construido máquinas semejantes. Entonces podemos preguntar si una computadora, que ha superado la prueba, *necesariamente* piensa, siente, comprende, etc. Volveré a este asunto más adelante. De momento, consideremos algunas de sus implicaciones. Por ejemplo, si los fabricantes tienen razón en sus afirmaciones más radicales, es decir, que su aparato es un ser pensante, sentimental, sensible, comprensivo, *consciente*, entonces la compra del aparato implicará *responsabilidades morales*. Esto realmente *debería* ser así si hemos de creer a los fabricantes. El simple hecho de poner en marcha la computadora para satisfacer nuestras necesidades sin tener en cuenta su propia sensibilidad, ya sería censurable. Sería lo mismo que maltratar a un esclavo. En general, tendríamos que evitar causar a la computadora el dolor que los fabricantes alegan que es capaz de sentir. Desconectar la computadora, o quizás incluso venderla cuando había llegado a sentirse muy unida a nosotros, nos plantearía dificultades morales, y habría otros incontables problemas del mismo tipo que se nos presentan en nuestra relación con otros seres humanos o con los animales. Todas estas cuestiones se volverían primordiales. Por todo eso sería para nosotros de gran importancia (y también para las autoridades) saber si las pretensiones de los fabricantes que, suponemos, se basan en su afirmación de que "cada uno de nuestros aparatos pensantes ha sido sometido a la prueba de Turing por un equipo de expertos" son realmente ciertas.

Creo que, pese al absurdo aparente de algunas de las implicaciones de este hecho, en particular las morales, el considerar la superación de la prueba de Turing como un indicio válido de la presencia de pensamiento, inteligencia, comprensión o conciencia *es* más que razonable, pues ¿de qué otro modo, si no es por la conversación juzgamos el que otras personas poseen tales cualidades? En realidad *sí existen* otros criterios, como las expresiones faciales, los movimientos corporales u otras acciones, que nos pueden influir de forma significativa al hacer tales juicios. Pero podemos imaginar que se pudiera construir (quizás en un futuro más lejano) un robot que imitase con éxito todas estas expresiones y movimientos. En ese caso ya no sería necesario

ocultar el robot y el sujeto humano de la vista de la interrogadora, aunque los criterios que ésta tendría a su disposición son, en principio, los mismos que antes.

En lo que a mí concierne, estoy dispuesto a relajar considerablemente los requisitos de la prueba de Turing. Creo que pedir a la computadora que imite a un ser humano de tal forma que resulte indistinguible de éste en los aspectos más importantes es, en verdad, pedirle más de la cuenta. Todo lo que yo pediría es que nuestra interrogadora perspicaz se sintiera realmente convencida —a través de la naturaleza de las réplicas de la computadora— de que hay una *presencia consciente*, aunque posiblemente extraña, que subyace en esas réplicas. Esto es algo manifiestamente ausente de todos los sistemas de computadoras que se han construido hasta la fecha. Me doy cuenta, sin embargo, de que existiría el peligro de que si la interrogadora fuera capaz de darse cuenta efectivamente de cuál de los sujetos era la computadora, entonces, quizás inconscientemente, podría ser reacia a atribuirle una conciencia, aun cuando *pudiera* percibirla. O, por el contrario, ella podría tener la impresión de que "siente" esa "presencia extraña" —y estar dispuesta a conceder a la computadora el beneficio de la duda— aun cuando no la hay. Por estas razones, la versión original de la prueba de Turing tiene una ventaja considerable al ser más objetiva y en general me atendré a ella en lo que sigue. La consiguiente "injusticia" que se comete con la computadora de la que he hablado antes (es decir, que para superar la prueba debe ser capaz de hacer todo lo que puede hacer un ser humano, mientras que el humano no necesita ser capaz de hacer todo lo que puede hacer una computadora) no es algo que parezca preocupar a los defensores de la prueba de Turing como una verdadera prueba de pensamiento. En cualquier caso, su reiterada opinión es que no pasará mucho tiempo antes de que una computadora pueda *realmente* superar la prueba, digamos hacia el año 2010. (Turing sugirió originalmente que para el año 2000 la computadora podría llegar al 30% de éxitos frente a un interrogador ' medio" y sólo cinco minutos de interrogatorio.) Sus partidarios parecen convencidos, en consecuencia, de que la falta de imparcialidad no está retrasando mucho ese día.

Todo esto resulta importante para una cuestión esencial: ¿realmente el punto de vista operacional proporciona un conjunto de criterios razonable para juzgar la presencia o la ausencia de cualidades mentales en un objeto? Algunos afirmarán contundentemente que no. La imitación, por muy hábil que sea, no es lo mismo que el objeto imitado. Mi posición a este respecto es en cierto modo intermedia. Como principio general me inclino a creer que la imitación, por muy hábil que sea, debería ser siempre detectable mediante un sondeo suficientemente hábil —aunque esto es más una cuestión de fe (o de optimismo científico) que un hecho probado. Por ello estoy dispuesto a aceptar la prueba como aproximadamente válida en su contexto. Es decir, *si* la computadora fuera capaz de responder a todas las preguntas que se le plantean de manera indistinguible a como lo haría un ser humano, y así, engañar completa* y consistentemente a nuestra interrogadora perspicaz, entonces, *en ausencia de cualquier evidencia en contra*, mi *conjetura* sería que la computadora realmente piensa y siente. Al utilizar palabras como "evidencia", "realmente", "conjetura", quiero decir que cuando me refiero a pensamiento, sentimiento o comprensión o, especialmente, a conciencia, considero que los conceptos significan "cosas" reales objetivas cuya presencia o ausencia en los cuerpos físicos es algo que tratamos de descubrir, y que no son simplemente conveniencias de lenguaje. Considero esto un punto crucial. Al tratar de discernir la presencia de tales cualidades hacemos conjeturas basadas en toda la evidencia disponible. (Esto no es diferente del caso, por ejemplo, de un astrónomo que trata de averiguar la masa de una estrella lejana.)

¿Qué tipo de evidencia en contra tendríamos que considerar? Es difícil establecer reglas por adelantado. No obstante, quiero dejar claro que el simple hecho de que la computadora pudiera estar construida a base de transistores y cables en lugar de neuronas y venas, *no* es, propiamente dicho, el tipo de cosas que consideraría como evidencias en contra. Estoy pensando en que en algún momento en el futuro pueda desarrollarse una teoría acertada de la conciencia —acertada en el sentido de que sea una teoría física coherente y apropiada, elegante y consistente con el resto de los conocimientos físicos, y tal que sus predicciones correspondan exactamente con las afirmaciones de los seres humanos acerca de cuándo o hasta qué punto parecen ellos mismos ser conscientes— y que esta teoría pueda tener implicaciones sobre la supuesta conciencia de nuestra computadora. Se podría incluso imaginar un "detector de conciencias", construido según los principios de esta teoría, que fuera completamente fiable frente a sujetos humanos pero que diera resultados diferentes a los de una prueba de Turing en el caso de una computadora. En tales circunstancias, tendríamos que ser muy cuidadosos a la * hora de interpretar los resultados de una prueba de Turing. Creo que ' la forma de ver la cuestión de cómo adaptar la prueba de Turing depende en parte de la forma en que esperamos que se desarrollen la ciencia y la tecnología. Tendremos que volver sobre estas consideraciones más adelante.

INTELIGENCIA ARTIFICIAL

Un área que ha despertado gran interés en los últimos años es la que se conoce como *inteligencia artificial*, a menudo abreviada simplemente como "IA". Los objetivos de la IA son imitar por medio de máquinas, normalmente electrónicas, tantas actividades mentales como sea posible, y quizá, llegar a mejorar las que llevan a cabo los seres humanos. El interés por los resultados de la IA procede al menos de cuatro direcciones. En concreto, tenemos el estudio de la *robótica*, que está interesada, sobre todo, en la aplicación industrial de los dispositivos mecánicos que pueden realizar tareas "inteligentes" —tareas de una variedad y complejidad que habían exigido anteriormente la intervención humana— y realizarlas con una velocidad y fíabilidad por encima de la de cualquier humano, o bien, en condiciones tales en las que la vida correría peligro. También es de interés comercial, así como general, el desarrollo de los llamados *sistemas expertos*, con los que se intenta codificar el conocimiento esencial de toda una profesión: medicina, abogacía, etc., en un paquete de ordenador. ¿Es posible que la experiencia y competencia de los profesionales pueda ser realmente reemplazada por estos paquetes? ¿O se trata sencillamente de que todo lo que podemos esperar son unas listas interminables de información objetiva y un sistema completo de referencias cruzadas? La cuestión de si las computadoras pueden mostrar (o imitar) inteligencia auténtica tiene evidentemente importantes implicaciones sociales.

Otra área en la que la IA podría tener importancia directa es la *psicología*: se confía en que tratando de imitar el comportamiento de un cerebro humano (o el de algún otro animal) mediante un dispositivo electrónico —o fracasando en el intento— podamos aprender cosas importantes sobre el funcionamiento cerebral. Finalmente, existe entre los optimistas la esperanza de que la

* Deliberadamente he permanecido cauto y sin revelar lo que considero sería una genuina aprobación de la prueba de Turing. Supongo, por ejemplo, que tras una larga serie de intentos fallidos de pasar la prueba, la computadora puede reunir todas las respuestas que el sujeto humano previamente le habría proporcionado y entonces simplemente devolverlas con ciertos adecuados ingredientes al azar. Después de un rato nuestro fatigado interrogador habrá agotado las preguntas originales que debía plantear y será timado de una manera que considero tramposa por parte de la computadora.

IA tuviera algo que decir sobre cuestiones profundas de la filosofía y que nos proporcionara algunos elementos nuevos del concepto *mente.*

¿Hasta dónde ha llegado la IA por el momento? Me resultaría difícil tratar de resumirlo. En diferentes partes del mundo existen muchos grupos activos y sólo estoy familiarizado con una pequeña parte de su trabajo. De todas formas, estaría bien decir que, aunque se han hecho muchas cosas ingeniosas, la simulación de algo que pudiera pasar por inteligencia auténtica tiene todavía un largo camino por delante. Para dar una, idea del tema mencionaré primero algunos de los logros anteriores (aún hoy en día impresionantes), y luego algunos progresos notables alcanzados recientemente con computadoras que juegan ajedrez.

Uno de los primeros dispositivos IA fue la "tortuga" de W. Grey Walter, construida a comienzos de los años cincuenta, [3] que se movía por el suelo hasta que sus baterías estaban bajas, entonces iba al enchufe más próximo, se conectaba y recargaba sus baterías. Una vez reabastecida se desconectaba por sí misma y volvía a sus aventuras por el suelo. Desde entonces se han construido muchos objetos similares (véase, por ejemplo, Waltz, 1982). Un ejemplo de desarrollo algo diferente es el programa de computadora de Terry Winograd, diseñado en 1972, que puede conversar (hablando y escuchando) sobre lo que hace con una colección de bloques de varias formas y colores que, por medio de simulación, coloca unos sobre otros en diferentes órdenes y disposiciones. Pese a estos primeros éxitos es notable cuan difícil ha resultado diseñar el control de un sencillo "brazo robot" articulado que hace maniobras evitando obstáculos amontonados, donde los movimientos necesarios parecerían "obvios" para una persona. La interpretación de imágenes visuales es también un área en la que los procedimientos de la IA no han conseguido acercarse a lo que el cerebro humano (y, de hecho, el de la mayoría de los animales) logra sin esfuerzo.

En línea más directa con la prueba de Turing está el programa del ordenador de K. M. Colby, desarrollado a mediados de los sesenta, que simula a un psicoterapeuta, con tanto éxito que algunos pacientes descubren que prefieren a la computadora antes que a un terapeuta humano, pues son más sinceros de lo que serían con su homólogo humano.

Como ejemplo [4] de tal intercambio entre la computadora "terapeuta" y un paciente, sigamos la siguiente conversación (mantenida a través de un teletipo):

TERAPEUTA: ¿En qué puedo ayudarle?

PACIENTE: No lo sé.

TERAPEUTA: Cuénteme algo.

PACIENTE: No.

TERAPEUTA: ¿Puede explicarse?

PACIENTE: Me siento enfermo y cansado.

TERAPEUTA: ¿Tiene alguna idea de por qué?

PACIENTE: Me siento viejo.

---

[3] Véase Gregory (1981), pp. 285-287, Grey Walter (1953).

[4] Este ejemplo se obtuvo de Delbrük (1986).

TERAPEUTA: Continúe, por favor.

PACIENTE: Creo que he perdido el interés en todo.

TERAPEUTA: Dígame algo más sobre eso.

PACIENTE: Tengo una idea vaga de por qué me siento así.

TERAPEUTA: ¿La tiene?

PACIENTE: Creo que quisiera tener una familia.'

TERAPEUTA: ¿Qué le hace desear una familia?

PACIENTE: Sería estimulante...

TERAPEUTA: ¿Por qué es tan importante para usted?

Aunque esto pueda dar una extraña impresión de que la computadora está entendiendo, en realidad no entiende nada; simplemente está siguiendo algunas reglas mecánicas sencillas. (Existe también un caso inverso a éste, en el que la computadora simula a un esquizofrénico, que presenta todos los síntomas y respuestas de un libro de texto y es *capaz* de engañar a algunos estudiantes de medicina que llegan a creer que realmente es un paciente humano el que está dando las respuestas.)

Las computadoras que juegan ajedrez proporcionan los mejores ejemplos de máquinas que poseen lo que podría ser considerado como "conducta inteligente". De hecho, algunas han alcanzado hoy día (1989) un nivel de juego más que respetable en relación con los jugadores humanos, acercándose al de "Maestro Internacional" (las puntuaciones de estas computadoras estarían por debajo de 2 300; en comparación con la del campeón mundial Kasparov, que está por encima de 2 700). En concreto, un programa de computadora (para el microprocesador comercial Fidelity Excel) de Dan y Kathe Spracklen ha alcanzado una puntuación (Elo) de 2110 y se le ha concedido el título de "Maestro" de la USCF. Aún más impresionante es "Deep Thought" (Pensamiento Profundo), programado fundamentalmente por Hsiung Hsu, de la Universidad de Carnegie Mellon, y que tiene una puntuación cercana a 2500 Elo, y recientemente logró la notable proeza de compartir el primer puesto (con el Gran Maestro Tony Miles) en un torneo de ajedrez (en Longbeach, California, en noviembre de 1988), derrotando por primera vez a un Gran [5] Maestro (Bent Larsen). Estas computadoras sobresalen también en la resolución de problemas de ajedrez y superan fácilmente a los humanos en este empeño.[6]

Las máquinas de jugar ajedrez dependen tanto del "conocimiento libresco" como de su poder de cálculo. Es digno de mención que, curiosamente estas máquinas en general son mejores en comparación con los jugadores humanos en el llamado "ajedrez-ping pong", cuando se impone que los movimientos se ejecuten muy rápidamente; en cambio, los jugadores humanos actúan

[5] Véase los artículos de O'Connell (1988) y Keene (1988). Para más información sobre las computadoras que juegan ajedrez, véase Levy (1984).

[6] Por supuesto la mayoría de los problemas de ajedrez están pensados para que sean difíciles para una *persona*. No sería demasiado complicado, creo, elaborar un problema que no fuera demasiado difícil para una persona pero que una computadora actual no pudiera resolver ni en mil años. (Habría que seguir un plan bastante obvio: plantear un problema cuya solución requiriera muchísimas jugadas. Son conocidos algunos que precisan más de 200, lo que es más que suficiente.) Esto propone un desafío interesante.

mucho mejor cuando se permite una buena cantidad de tiempo para cada movimiento. Esto ha de ser debido a que las decisiones de la computadora se basan en extensos cálculos rápidos y exactos, mientras que el jugador humano saca ventaja de consideraciones conscientes relativamente lentas. Los juicios humanos reducen drásticamente el número de posibilidades que deben considerarse seriamente en cada etapa del cálculo, y cuando *se dispone* de tiempo se puede hacer un análisis mucho más profundo que el del mero cálculo y la eliminación directa de posibilidades, como lo hace la máquina. (Esta diferencia es aún más notable en el "go", difícil juego oriental, en el que el número de posibilidades en cada movimiento es mucho mayor que en el ajedrez.) La relación entre conciencia y formación de juicios será capital para mis argumentos posteriores, especialmente en el capítulo X.

LA APROXIMACIÓN DE LA IA AL "PLACER" Y AL "DOLOR"

Una de las pretensiones de la IA es proporcionar una vía hacia el entendimiento de las cualidades mentales, tales como la felicidad, el dolor o el hambre. Tomemos por ejemplo la tortuga de Grey Walter. Cuando sus baterías estén bajas su pauta de comportamiento cambiará y actuará de la forma planeada para reabastecer su reserva de energía. Existen claras analogías entre ésta y la manera en que actuaría un ser humano —o cualquier otro animal— cuando sienta hambre. No sería un grave abuso de lenguaje decir que la tortuga de Grey Walter está "hambrienta" cuando actúa de esta forma. Algún mecanismo interno es sensible al estado de carga de su batería, y cuando éste caía por debajo de cierto nivel, orientaba a la tortuga hacia una pauta de comportamiento diferente. Sin duda existe una operación similar en los animales cuando empiezan a tener hambre, sólo que los cambios de comportamiento son más complicados y sutiles. Más que pasar de una pauta de comportamiento a otra, hay un cambio en las *tendencias* a actuar de cierta forma, estos cambios son tanto más fuertes (hasta cierto punto) en la medida en que aumenta la necesidad de reabastecerse de energía.

De modo análogo, los defensores de la IA imaginan que conceptos tales como el dolor o la felicidad pueden modelarse adecuadamente de esta forma. Simplifiquemos las cosas y consideremos sólo una escala de sentimientos" que va desde el "dolor" extremo (puntuación -100) al "placer" extremo (puntuación +100). Imaginemos que tenemos un dispositivo —una máquina de algún tipo, presumiblemente electrónica— que tiene algún medio de registrar su propia (supuesta) puntuación "placer-dolor", que llamaré "puntuación-pd". El dispositivo tiene ciertas formas de comportamiento y ciertos datos de entrada, ya sean internos (como el estado de sus baterías) o externos. La idea es que sus acciones estén ajustadas para conseguir la máxima puntuación-pd. Habría muchos factores que influirían en la puntuación-pd. Podríamos ciertamente disponer que la carga de sus baterías fuera uno de ellos, de modo que una carga baja contara negativamente y una carga alta positivamente, pero habría otros factores. Quizá nuestro dispositivo tuviera algunos paneles solares que le proporcionaran medios alternativos de obtener energía, de modo que no fuera necesario hacer uso de sus baterías cuando los paneles estuvieran en operación. Podríamos disponer que al moverse hacia la luz incrementara algo su puntuación-pd, de modo que, en ausencia de otros factores, eso sería lo que tendería a hacer. (En realidad, la tortuga de Grey Walter acostumbraba a *evitar* la luz.) Sería necesario tener algún medio de realizar cálculos para que pudiese evaluar los probables efectos que sus diferentes acciones tendrían en su puntuación-pd. Podrían introducirse pesos relativos, de modo que un cálculo

tuviera un efecto mayor o menor en la puntuación dependiendo de la confiabilidad de los datos en los que se basaba.

También sería necesario proporcionar a nuestro dispositivo objetivos diferentes a los del simple mantenimiento del suministro de energía, ya que de lo contrario no tendríamos modo de distinguir el "dolor" del "hambre". Sin duda es demasiado pedir que nuestro dispositivo tenga medios de procreación así que, de momento, ¡nada de sexo! Pero quizá podamos implantar en él el "deseo" de compañía de otro dispositivo semejante, si damos a sus encuentros una puntuación-pd positiva. O podríamos hacer que estuviera "ansioso" de aprender simplemente por gusto, de modo que el almacenamiento de datos sobre el mundo externo tuviera también puntuación positiva en su escala-pd. (Más egoístamente, podríamos disponer que al realizar para *nosotros* diferentes servicios tuviera una puntuación positiva, como tendríamos que hacer si construyéramos un criado robot.) Podría alegarse que hay algo artificioso en imponer a nuestro capricho tales objetivos al dispositivo; no obstante, esto no es muy diferente al modo en que la selección natural nos ha impuesto, como individuos, ciertos "objetivos" que están gobernados en gran medida por la necesidad de propagar nuestros genes.

Supongamos ahora que nuestro dispositivo ha sido construido con éxito de acuerdo con todo lo anterior. ¿Qué tan objetivos seríamos al asegurar que realmente *siente* placer cuando su puntuación-pd es positiva y dolor cuando la puntuación es negativa? El punto de vista conductista de la IA diría que juzguemos esto simplemente a partir del modo en que se comporta. Puesto que actúa de una forma que incrementa su puntuación tanto como sea posible (y durante tanto tiempo como sea posible), y como también actúa para evitar puntuaciones negativas, entonces podemos *definir* razonablemente su sentimiento de placer como el "grado de positividad" de su puntuación, y *definir* su sentimiento de dolor como el "grado de negatividad" de esa puntuación. La coherencia de tal definición, se alegaría, deriva del hecho de que ésta es precisamente la forma en que reacciona un ser humano en relación con los sentimientos de placer o dolor. Por supuesto que con los seres humanos las cosas en realidad no son tan sencillas, como ya sabemos, a veces parecemos buscar el dolor deliberadamente o apartarnos de nuestro camino para evitar ciertos placeres. Es evidente que nuestras acciones están realmente guiadas por criterios mucho más complejos que éstos (*cfr.* Dennett, 1979, pp. 190-229). Pero, en primera instancia, nuestra forma de actuar consiste en evitar el dolor y buscar el placer. Para un conductista esto sería suficiente para considerar la *identificación* de la puntuación-pd de nuestro dispositivo con su valoración placer-dolor. Tales identificaciones parecen estar también entre los propósitos de la teoría de la IA.

Debemos preguntar: ¿es realmente cierto que nuestro dispositivo *siente* dolor cuando su puntuación-pd es negativa y placer cuando es positiva? De hecho, ¿podría nuestro dispositivo sentir, a secas? Sin duda, los operacionalistas dirán que obviamente sí, o tacharán tales preguntas de absurdas. A mí, en cambio, me parece evidente que *existe* una cuestión seria y difícil que debe ser considerada. Las influencias que nos impulsan a nosotros mismos son de varios tipos. Algunas son conscientes, como el dolor y el placer, pero hay otras de las que no tenemos conciencia directa. Esto queda claramente ilustrado en el ejemplo de una persona que toca una estufa caliente. Tiene lugar una acción involuntaria y retira la mano aun antes de que experimente cualquier sensación de dolor. Parecería que tales acciones involuntarias están mucho más cerca de las respuestas de nuestro dispositivo a su puntuación-pd que de los efectos reales del dolor y el placer.

Con frecuencia utilizamos términos antropomorfos, en forma descriptiva, a menudo jocosa, para referirnos al comportamiento de las máquinas: "Mi coche no quería arrancar esta mañana", o "mi reloj aún piensa que va con la hora de California", o "mi computadora afirma que no entiende la última instrucción y que no sabe cómo continuar". Por supuesto, no queremos decir que el coche realmente *quiere* algo, que el reloj *piensa*, que la computadora* *afirma* algo o *comprende* o incluso que *sabe* lo que está haciendo. De todas formas, tales proposiciones pueden ser meramente descriptivas y útiles para nuestra comprensión del tema, con tal de que las tomemos simplemente en el sentido con el que fueron pronunciadas y no las consideremos como aserciones literales. Adoptaré una actitud similar respecto a las diversas afirmaciones de la IA acerca de las cualidades mentales que podrían estar presentes en los dispositivos que hemos estado construyendo, *independientemente* del ánimo con que se planearon. Si acepto que se diga que la tortuga de Grey Walter está hambrienta, es precisamente en este sentido medio gracioso. Si estoy dispuesto a utilizar términos como "dolor" o "placer" para la puntuación-pd de un dispositivo como se concibió más arriba, es porque encuentro estos términos útiles para la comprensión de su comportamiento, debido a ciertas analogías con mi propia conducta y estados mentales. No quiero decir que estas analogías sean particularmente estrechas o que no haya otras cosas inconscientes que influyan en mi comportamiento de forma mucho *más* parecida.

Confío en que para el lector quede claro que, en mi opinión, hay mucho más que entender de las cualidades mentales de lo que puede obtenerse directamente de la IA. De todas formas, creo que la IA plantea un abordaje serio que debe ser respetado y sometido a consideración. Con esto no quiero decir que se haya conseguido mucho —si es que se ha conseguido algo— en la simulación de la inteligencia real. Pero hay que tener en cuenta que la disciplina es muy joven. Próximamente las computadoras serán más rápidas, tendrán mayores memorias de acceso rápido, más unidades lógicas y podrán realizar un mayor número de operaciones en paralelo. Habrá progresos en el diseño lógico y en la técnica de programación. Estas máquinas, portadoras de la filosofía de la IA, serán enormemente perfeccionadas en sus atributos técnicos. Además, la filosofía misma *no* es intrínsecamente absurda. Quizá la inteligencia humana pueda ser aproximadamente simulada por las computadoras electrónicas, esencialmente las actuales, basadas en principios que ya son comprendidos, pero que en los próximos años tendrán capacidad, velocidad, etc., mucho mayores. Quizá, incluso, estos dispositivos *serán* realmente inteligentes; quizá pensarán, sentirán y tendrán una mente. O quizá no, y se necesite algún principio nuevo del que por el momento no hay indicios. Esto es lo que está en discusión, y es algo que no puede despacharse a la ligera. Así que trataré de presentar evidencias, de la mejor manera posible, de mis propias ideas.

LA IA FUERTE LA HABITACIÓN CHINA DE SEARLE

Hay un punto de vista conocido como el de la *IA fuerte* que adopta una posición más bien extrema sobre estas cuestiones.[7] Según la IA fuerte, los dispositivos que acabamos de mencionar no sólo son inteligentes y tienen una mente, sino que al funcionamiento lógico de *cualquier* dispositivo computacional se le puede atribuir un cierto tipo de cualidades mentales, incluso los

* Al menos como las de 1989.

7 A lo largo de este libro he adoptado la terminología de Searle "IA fuerte" para designar este punto de vista extremo, sólo por ser concreto. El término "funcionalismo" se utiliza frecuentemente para representar la misma idea, pero quizá no siempre de una forma tan concreta. Sostienen este punto de vista Minsky (1968), Fodor (1983), Hofstadter (1979) y Moravec(1989).

dispositivos mecánicos más simples, como un termostato.[8] La idea es que la actividad mental consiste simplemente en una secuencia bien definida de operaciones, frecuentemente llamada *algoritmo.* Más adelante precisaré lo que realmente es un algoritmo. Por el momento, será suficiente definir un algoritmo como cierto tipo de procedimiento de cálculo. En el caso del termostato el algoritmo es extremadamente simple: el dispositivo registra si la temperatura es mayor o menor que la establecida, y a continuación dispone que el circuito se desconecte o se conecte, según el caso. Para cualquier tipo importante de actividad mental en el cerebro humano el algoritmo tendría que ser muchísimo más complicado pero, según el punto de vista de la IA fuerte, un algoritmo complejo diferirá enormemente sólo en el grado del sencillo algoritmo del termostato, pero no habrá diferencia de principio. Así, según la IA fuerte, la diferencia entre el funcionamiento esencial del cerebro humano (incluyendo todas sus manifestaciones conscientes) y el de un termostato radica sólo en que el primero posee una mucho mayor *complicación* (o quizá "mayor orden de estructura" o "propiedades auto-referentes", u otro atributo que pudiéramos asignar a un algoritmo). Y lo que es más importante, todas las cualidades mentales —pensamiento, sentimiento, inteligencia, comprensión, conciencia— deben ser consideradas, según este punto de vista, simplemente como aspectos de este funcionamiento complicado; es decir, son simplemente características del *algoritmo* que ejecuta el cerebro.

La virtud de cualquier algoritmo específico reside en su desempeño, es decir, en la precisión de sus resultados, su amplitud, su economía y la velocidad con que puede ser ejecutado. Un algoritmo que pretenda igualar el que se presume está operando en el cerebro humano tendría que ser algo prodigioso. Pero si existiera un algoritmo de esta especie para el cerebro —y los defensores de la IA fuerte afirmarán ciertamente que sí existe— entonces podría en principio funcionar en una computadora.

De hecho podría funcionar en *cualquier* computadora electrónica moderna de tipo general si no fuera por limitaciones de espacio de almacenamiento y velocidad de operación. (La justificación de este comentario vendrá más tarde, cuando consideremos la máquina universal de Turing.) Se prevé que cualquiera de estas limitaciones habrá quedado superada en las grandes y rápidas computadoras de un futuro no muy lejano. En esa eventualidad, un algoritmo así superaría presumiblemente la prueba de Turing. Los defensores de la IA fuerte alegarán que, donde quiera que funcione, el algoritmo experimentará *autónomamente* sentimientos y tendrá una conciencia. Será la mente.

No todo el mundo estará de acuerdo, ni mucho menos, en que se puedan identificar así los estados mentales con los algoritmos. En particular, el filósofo americano John Searle (1980, 1987) se ha opuesto enérgicamente a esta idea. Cita ejemplos de versiones simplificadas de la prueba de Turing que han sido superadas *efectivamente* por una computadora adecuadamente programada, pero da argumentos de peso de que el atributo mental de la "comprensión" está totalmente ausente. Uno de estos ejemplos se basa en el programa de ordenador diseñado por Roger Schank (Schank y Abelson, 1977). El propósito del programa es simular la comprensión de historias sencillas como: "Un nombre entró en un restaurante y pidió una hamburguesa. Cuando se la trajeron estaba quemada y el hombre salió vociferando furiosamente del restaurante, sin pagar la cuenta ni dejar propina." Un segundo ejemplo: "Un hombre entró en un restaurante y pidió una hamburguesa; cuando se la trajeron, le gustó mucho, y al salir del restaurante dio una buena propina al camarero antes de pagar su cuenta." En la prueba de

[8] Véase Searle (1987), p. 211, para un ejemplo de tal afirmación.

"comprensión" de las historias se le pregunta a la computadora si el hombre comió la hamburguesa en cada caso (algo que no se había mencionado explícitamente en ninguna de las dos historias). Para este tipo de historia y preguntas sencillas la computadora puede dar respuestas que son esencialmente indistinguibles de las respuestas que daría un ser humano de habla española: "no", en el primer caso, "sí", en el segundo. En este caso, en este sentido *muy* limitado, la máquina ya ha superado una prueba de Turing. Ahora bien, la cuestión que debemos considerar es si un acierto de este tipo es indicio realmente de una auténtica comprensión por parte de la computadora o, quizá, por parte del propio programa. El argumento de Searle en contra se expresa en su "habitación china". En primer lugar, imagina que las historias son contadas en chino y no en inglés —ciertamente un cambio que no es esencial— y que todas las operaciones del algoritmo de la computadora para este ejercicio concreto se suministran (en inglés) como un conjunto de instrucciones para manipular fichas con símbolos chinos en ellas. Searle se imagina *a sí mismo* haciendo todas las manipulaciones en el interior de una habitación cerrada. Las secuencias de símbolos que representan primero las historias, y luego las preguntas, se introducen en la habitación a través de una pequeña ranura. No se permite ninguna otra información del exterior. Finalmente, cuando se han completado todas las manipulaciones, la secuencia resultante se entrega a través de la ranura. Puesto que todas estas manipulaciones simplemente ejecutan el algoritmo de Schank, el resultado final será simplemente el equivalente chino de "sí" o "no", según sea el caso, con las que se responderá a una pregunta formulada en chino acerca de la historia también en chino. Searle deja en claro que él no entiende una sola palabra de chino, de modo que no tiene la más remota idea de lo que cuentan las historias. De todas formas, llevando a cabo correctamente la serie de operaciones que constituyen el algoritmo de Schank (las instrucciones para este algoritmo le han sido dadas en inglés) sería capaz de contestar tan bien como lo haría una persona china que realmente entendiera las historias. El punto importante de Searle —y pienso que es bastante convincente— es que la mera ejecución de un algoritmo correcto *no* implica en sí mismo que haya tenido lugar comprensión alguna. El imaginario Searle, encerrado en su habitación china, no comprenderá ni jota de ninguna de las historias.

Se han levantado diversas objeciones contra el argumento de Searle. Mencionaré sólo las que considero de mayor importancia. En primer lugar, hay algo más bien impreciso en la frase "no comprenderá ni jota" que se utilizó antes. Ya que la comprensión tiene tanto que ver con estructuras como con las letras y las palabras aisladas. Mientras se ejecutan algoritmos de este tipo, se podría perfectamente empezar a percibir algo de la estructura que forman los símbolos sin comprender realmente el significado de muchos de ellos individualmente. Por ejemplo, el carácter chino de "hamburguesa" (si es que existe tal cosa) podría ser reemplazado por el de cualquier otro plato, pongamos por caso "chou mein", y las historias no se verían seriamente afectadas. De todas formas, me parece razonable suponer que (incluso considerando que tales sustituciones sean importantes) muy pocos de los significados reales de las historias se concretarían si simplemente se continuasen ejecutando los detalles de tal algoritmo.

En segundo lugar, debemos tener en cuenta el hecho de que incluso la ejecución de un programa bastante sencillo podría ser extraordinariamente larga y tediosa si fuera realizada por seres humanos manipulando símbolos. (Después de todo, por eso tenemos computadoras que hacen esas cosas para nosotros.) Si Searle tuviera que ejecutar realmente el algoritmo de Schank de la forma sugerida, necesitaría probablemente muchos días, meses o años de trabajo extremadamente pesado para responder sólo a una sencilla pregunta (una actividad no del todo

indicada para un filósofo). Sin embargo, ésta no me parece una objeción seria puesto que aquí estamos interesados en cuestiones *de principio y* no en cuestiones prácticas. La dificultad es mayor si el supuesto programa tiene suficiente complicación como para igualar al cerebro humano y, por lo tanto, superar *perfectamente* la prueba de Turing. Un programa semejante sería terriblemente complicado. Podemos imaginar que la ejecución de este programa, para dar respuesta a alguna pregunta de la prueba de Turing, por simple que fuera, podría suponer tantos pasos que no habría posibilidad de que ningún ser humano llevara a cabo manualmente el algoritmo, aunque le dedicara toda su vida. En ausencia de tal programa es difícil decir si realmente será así.[9] Pero en cualquier caso, la cuestión de la extrema complejidad no puede, en mi opinión, ser simplemente pasada por alto. Es cierto que aquí estamos interesados en cuestiones de principio, pero aún suponiendo que pudiera haber algún grado "crítico" de complejidad para que un algoritmo muestre cualidades mentales, este valor crítico debería ser tan grande y hasta tal punto complejo que sería inconcebible que fuera ejecutado a mano por ningún ser humano de la manera imaginada por Searle.

El propio Searle ha replicado a esta última objeción permitiendo que todo un equipo de seres humanos manipuladores de símbolos y que no hablen el chino reemplace al anterior habitante único (él mismo) de su habitación china. Para tener una cantidad considerable imagina su habitación reemplazada por toda la India, con toda su población (excluyendo los que entienden chino, claro está) ocupados ahora en la manipulación de símbolos. Aunque esto sería absurdo en la práctica, no lo es *en principio*, *y* el argumento es esencialmente el mismo que antes: los manipuladores de símbolos *no* comprenden la historia, pese a la afirmación de la IA fuerte de que la simple ejecución del algoritmo apropiado daría lugar a la cualidad mental de "comprensión". Sin embargo, otra cuestión comienza ahora a cobrar importancia. ¿No es cada uno de los hindúes más semejante a las neuronas individuales del cerebro de una persona que al cerebro global? Nadie sugerirá que las neuronas, cuyas descargas constituyen aparentemente la actividad física del cerebro en el acto de pensar, comprendan *individualmente* lo que la persona está pensando, así que ¿por qué esperar que los hindúes como individuos comprendan las historias chinas? Searle replica a esta sugerencia señalando el aparente absurdo de imaginar a la India, como país, comprendiendo una historia que ninguno de sus habitantes individuales comprende. Un país, aduce, igual que un termostato o un automóvil, no está dedicado al "negocio de la comprensión" mientras que una persona individual sí.

Este argumento tiene mucho menos fuerza que el anterior. En mi opinión, el argumento de Searle adquiere su mayor fuerza cuando hay un solo individuo ejecutando el algoritmo que suponemos suficientemente sencillo para que una persona lo ejecute en un tiempo inferior a la duración de la vida humana. No considero que este argumento establezca *rigurosamente* que no habría ningún tipo de "comprensión" incorpórea asociada con las personas que ejecutan el algoritmo, y cuya presencia no chocaría con sus propias conciencias. Sin embargo, coincido con Searle en que esta posibilidad es, cuando menos, bastante remota. Sin embargo, el argumento de Searle tiene una fuerza considerable aunque no sea del todo concluyente. Es convincente sobre todo al demostrar que algoritmos con el tipo de complejidad del programa de Schank no pueden tener verdadera

[9] En su crítica al artículo original de Searle, tal como está reimpresa en *The Mind's I*, Douglas Hofstadter considera inconcebible que algún ser humano podría "absorber" la descripción completa de la mente de otro ser humano, debido a la complicación que supone. Claro que no, pero, tal como yo lo veo, no es esa la idea exacta. Estamos interesados simplemente en ejecutar parte de un algoritmo que pretende representar lo que ocurre en un estado mental particular. Este podría ser cualquier "comprensión consciente" momentánea al contestar una pregunta en una prueba de Turing, o podría incluso ser algo más simple. ¿Realmente requeriría tal suceso un algoritmo de enorme complicación?

comprensión de ninguna de las tareas que ejecutan; también *sugiere* (aunque sólo eso) que ningún algoritmo, por muy complejo que sea, podrá, por sí mismo, desarrollar nunca auténtica comprensión, en contraste con las tesis de la IA fuerte.

Existen, me parece, otras dificultades muy serias en el punto de vista de la IA fuerte. Según sus teóricos, lo que cuenta es simplemente el algoritmo. No hay ninguna diferencia si el algoritmo es ejecutado por un cerebro, una computadora electrónica, una nación entera de hindúes, un dispositivo mecánico de ruedas y engranajes o un sistema de tuberías. La idea reside en que es simplemente la estructura lógica del algoritmo lo significativo del "estado mental" que se supone representa, siendo completamente irrelevante la encarnación física de dicho algoritmo. Como apunta Searle, esto entraña de hecho una forma de "dualismo". El *dualismo* es un punto de vista filosófico adoptado por el muy influyente filósofo y matemático del siglo XVII René Descartes, y afirma que hay dos tipos de substancias distintas: "substancia mental" y materia ordinaria. El que uno de estos tipos de substancia pueda o no afectar al otro, o de qué modo pueda hacerlo, es una cuestión adicional. El punto importante es que se supone que la substancia mental no está compuesta de materia y puede existir independientemente de ella. La substancia mental de la IA fuerte es la estructura lógica de un algoritmo. Como acabo de señalar, la encarnación física concreta de un algoritmo es algo totalmente irrelevante. El algoritmo tiene un tipo de "existencia" incorpórea que es ajena a cualquier realización de dicho algoritmo en términos físicos. Hasta qué punto debemos considerar seriamente este tipo de existencia es una cuestión sobre la que volveré en el siguiente capítulo: es parte del problema general de la realidad platónica de los objetos matemáticos abstractos. Por el momento dejaré de lado este tema general e indicaré simplemente que los defensores de la IA fuerte parecen estar tomando seriamente la realidad de los algoritmos, ya que creen que los algoritmos forman la substancia de sus pensamientos, sus sentimientos, su entendimiento y sus percepciones conscientes. Resulta curiosamente irónico, como Searle ha señalado, el hecho de que el punto de vista de la IA fuerte parezca llevarnos a una forma extrema de dualismo: precisamente el punto de vista con el que menos desearían estar asociados sus defensores.

El dilema se halla entre los bastidores de un argumento desarrollado por Douglas Hofstadter (1981) —un defensor importante de la IA fuerte— en un diálogo titulado "Conversación con el cerebro de Einstein". Hofstadter imagina un libro, de proporciones absurdamente monstruosas, que se supone contiene una completa descripción del cerebro de Albert Einstein. Cualquier pregunta que uno pudiera plantear a Einstein podría ser respondida, exactamente igual que lo hubiera hecho Einstein en vida, simplemente hojeando el libro y siguiendo cuidadosamente todas las instrucciones detalladas que proporciona. Por supuesto, "simplemente" es una palabra totalmente inadecuada, como Hofstadter se cuida en señalar. Pero su tesis es que *en principio* el libro es exactamente equivalente, en el sentido operacional de la prueba de Turing, a una versión ridículamente disminuida del Einstein real. Así, según las opiniones de la IA fuerte, el libro pensaría, sentiría, entendería, sería consciente, exactamente como si fuera el propio Einstein, pero viviendo en cámara lenta (de modo que para el Einstein libro el mundo externo parecería discurrir como una exhalación a un ritmo ridículamente acelerado). De hecho, ya que se supone que el libro es simplemente una particular encarnación del algoritmo que constituía el "yo" de Einstein, realmente *sería* Einstein.

Pero ahora se presenta una nueva dificultad. El libro podría no ser abierto nunca o, por el contrario, podría ser examinado continuamente por innumerables estudiantes e investigadores en pos de la verdad. ¿Cómo sabría el libro la diferencia? Tal vez el libro no necesitara ser abierto si

su información fuera recuperada mediante rayos X, tomografía o cualquier otro prodigio técnico. ¿Estaría activa la conciencia de Einstein sólo cuando el libro esté siendo examinado de esta forma? ¿Sería dos veces consciente si dos personas decidiesen plantearle la misma pregunta en dos momentos diferentes? ¿O ello implicaría dos casos separados y totalmente diferentes del *mismo* estado de conciencia de Einstein?

Quizá esta conciencia se activa sólo si se *altera* el libro? Después de todo, cuando somos conscientes de algo recibimos información del mundo externo que afecta a nuestra memoria, y el estado de nuestra mente cambia ligeramente. Si es así, ¿significa esto que son los *cambios* (apropiados) en los algoritmos (y aquí incluyo la memoria de almacenamiento como parte del algoritmo) los que deben ser asociados con sucesos mentales en lugar de (o quizás además de) la *ejecución* de los algoritmos? ¿Permanecería el Einstein-libro completamente autoconsciente aún si nunca fuera examinado o perturbado por nadie o nada? Hofstadter toca alguna de estas cuestiones, pero no intenta responderlas ni llegar a conclusión alguna en la mayoría de ellas.

¿Qué significa ejecutar un algoritmo, o encarnarlo en forma física? ¿Habría alguna diferencia entre cambiar un algoritmo y simplemente descartarlo y reemplazarlo por otro? ¿Qué demonios tiene todo esto que ver con nuestras sensaciones de conciencia? El lector (a menos que sea defensor de la IA fuerte) puede estar preguntándose por qué he dedicado tanto espacio a una idea tan evidentemente absurda. En realidad, yo *no* considero la idea absurda en sí, sino especialmente errónea. De hecho, hay que reconocer cierta fuerza al razonamiento que sustenta la IA fuerte, y esto es lo que trataré de explicar. En mi opinión, hay también un cierto atractivo en algunas de sus ideas —si se modifican adecuadamente— como también trataré de mostrar. Además, considero que el punto de vista contrario expresado por Searle también implica serias dificultades y absurdos aparentes, pero aun así estoy en buena parte de acuerdo con él.

Searle, en su exposición, parece aceptar implícitamente que las computadoras electrónicas del tipo de las actuales, pero con una velocidad de acción y una memoria de acceso rápido considerablemente aumentadas (y posiblemente con acción paralela), podrían ser perfectamente capaces de pasar limpiamente la prueba de Turing en un futuro no muy lejano. Está dispuesto a aceptar la idea de la IA fuerte (y de muchos otros puntos de vista "científicos") de que "somos una materialización de ciertos programas de cómputo". Además, cede y afirma: "Por supuesto el cerebro es una computadora digital. Puesto que cualquier cosa es una computadora digital, los cerebros también lo son.[10] Searle sostiene que la diferencia entre la función del cerebro humano (que puede alojar a la mente) y la de las computadoras electrónicas (que, según él, no pueden hacerlo), pudiendo ambas ejecutar un mismo algoritmo, radica solamente en la construcción material de cada uno. Dice, aunque no es capaz de explicar las razones, que los objetos biológicos (cerebros) pueden poseer "intencionalidad" y "semántica", lo que él considera como las características definitorias de la actividad mental, mientras que los electrónicos no. En sí mismo esto no creo que señale ningún camino hacia una teoría de la mente científicamente útil. ¿Qué hay tan especial en los sistemas biológicos —aparte quizá de la forma "histórica" en que han evolucionado (y el hecho de que *nosotros* seamos uno de esos sistemas)__,que los discrimina como los únicos objetos a los que se permite alcanzar intencionalidad o semántica? La tesis me parece sospechosamente dogmática, quizá no menos dogmática, incluso, que las afirmaciones de la IA fuerte que sostienen que la simple ejecución de un algoritmo puede producir un estado de conciencia.

[10] Véanse pp. 368, 372 del artículo de Searle (1980) en Hofstadter y Dennett (1981).

En mi opinión Searle, y muchas otras personas, han sido confundidas por los computólogos e informáticos. Y ellos, a su vez, han sido confundidos por los físicos. (No es culpa de los físicos. Ni siquiera *ellos* lo saben todo.) Parece ser una creencia muy extendida el que cualquier cosa es una computadora digital. Mi intención en este libro es tratar de demostrar por qué, y quizá cómo, esto *no* tiene que ser así necesariamente.

HARDWARE Y SOFTWARE

En la jerga de la computación se utiliza el término *hardware* para designar la maquinaria real de un ordenador (circuitos impresos, transistores, cables, memorias magnéticas, etc.), incluyendo la descripción detallada del modo en que todo está interconectado. Asimismo, el término *software se* refiere a los diversos programas que pueden ser ejecutados en la máquina. Uno de los descubrimientos notables de Turing fue el de que, en realidad, cualquier máquina para la que el *hardware* ha alcanzado un cierto grado de complejidad y flexibilidad, es *equivalente* a cualquier otra máquina semejante. Esta equivalencia debe tomarse en el sentido de que para dos de estas máquinas, A y B, existirá un *software* específico tal que si se le proporcionara a la máquina A, ésta actuaría exactamente como si fuera la máquina B; del mismo modo, existiría otro *software* que haría que la máquina B actuara exactamente como la máquina A. Utilizo aquí la palabra "exactamente" en referencia a la respuesta (*output*) real de las máquinas para cualquier estímulo (*input*) dado (realizado después de que se haya instalado el *software* convertidor) y *no* al *tiempo* que podría tardar cada máquina en producir esa respuesta. Por supuesto, si cualquiera de las máquinas agota en un momento dado el espacio de almacenamiento para sus cálculos, puede recurrir a algún suministro externo (en principio ilimitado) de "papel" en blanco, que podría tomar la forma de cinta magnética, discos, bobinas o cualquier otra cosa.

Ciertamente, la diferencia en el tiempo que necesitan las máquinas A y B para ejecutar alguna tarea puede ser una consideración importante. Pudiera ser, por ejemplo, que A fuera mil veces más rápida que B en realizar una tarea particular. También podría darse el caso de que, para las mismas máquinas, hubiera otra tarea para la que B es mil veces más rápida que A. Además, estos tiempos podrían depender en buena medida del *software* convertidor que se haya elegido. Evidentemente esta es una discusión "de principios", en la cual no estamos interesados realmente en las cuestiones prácticas como el realizar los cálculos en un tiempo razonable. En la próxima sección precisaré más los conceptos que utilizo aquí; las máquinas A y B son ejemplos de las llamadas *máquinas universales de Turing*.

En realidad, todas las computadoras modernas de uso general son máquinas universales de Turing. Por lo tanto, todas ellas son equivalentes en el sentido anterior. Las diferencias entre sí pueden ser reducidas al *software*, siempre que no estemos interesados en la velocidad de operación resultante ni en las posibles limitaciones de memoria. De hecho, la tecnología moderna ha posibilitado que las computadoras actúen con tanta velocidad y con tan grandes capacidades de almacenamiento que, para la mayoría de los usos cotidianos, ninguna de estas consideraciones prácticas representa una limitación seria para lo que se requiere normalmente.[*] Así pues, esta equivalencia teórica entre computadoras puede verse también en el nivel práctico. La tecnología ha transformado —por lo menos así lo parece— las discusiones estrictamente

[*]. Sin embargo, véase la discusión acerca de la teoría de la complejidad y los problemas NP al final del capítulo IV.

académicas acerca de los conceptos teóricos de cómputo en cuestiones que afectan directamente nuestra vida cotidiana.

Me parece que uno de los factores mas interesantes que sustenta la filosofía de la IA fuerte es esta equivalencia entre dispositivos físicos de cómputo. El *hardware* se considera relativamente de poca importancia (quizá incluso totalmente irrelevante) y, en cambio, el *software*, esto es, el programa o los algoritmos, es un ingrediente vital. Sin embargo, creo que en el fondo hay otros factores importantes subyacentes que se originan en la física. Trataré de indicar cuáles son estos factores.

¿Qué es lo que da su identidad al individuo? ¿Son, de alguna manera, los propios átomos que componen su cuerpo? ¿Depende su identidad de la elección particular de electrones, protones y otras partículas que componen estos átomos? Hay al menos dos razones por las que no puede ser así. En primer lugar, hay una continua renovación en el material del cuerpo de cualquier persona viva, esto se aplica también a las células del cerebro a pesar de que no se producen células cerebrales después del nacimiento. La inmensa mayoría de los átomos en cada célula viva (incluyendo cada célula del cerebro) y, de hecho, virtualmente todo el material de nuestros cuerpos, han sido reemplazados varias veces desde el nacimiento.

La segunda *razón* procede de la mecánica cuántica —y por una extraña ironía está, estrictamente hablando, en contradicción con la primera. Según ésta (y veremos más sobre esto en el capítulo VI) dos electrones cualesquiera deben ser por necesidad totalmente idénticos, y lo mismo sucede para dos protones y para dos partículas cualesquiera de cualquier tipo específico. Esto no quiere decir que no haya manera de distinguir las partículas: el enunciado es mucho más profundo que eso. Si se intercambiara un electrón del cerebro de una persona con un electrón de un ladrillo, el estado del sistema sería *exactamente*[11] *el mismo que antes*, y no simplemente indistinguible de él. Lo mismo sucede para protones o cualquier otro tipo de partícula, y para átomos enteros, moléculas, etc. Si todo el contenido material de una persona fuera intercambiado con las correspondientes partículas de los ladrillos de su casa entonces, en un sentido general, no habría sucedido nada en absoluto. Lo que diferencia a la persona de su casa es la *pauta* con que están dispuestos sus constituyentes, y no la individualidad de esos constituyentes.

En el ámbito cotidiano hay algo análogo y que es independiente de la mecánica cuántica, pero que se me hizo plenamente manifiesto cuando escribía esto gracias a la tecnología electrónica que me permite teclear en un procesador de textos. Si quiero cambiar una palabra, transformar, pongamos por caso, "casa" en "cosa", puedo hacerlo simplemente reemplazando la "a" por una "o", o puedo decidir en su lugar teclear de nuevo toda la palabra. Si hago esto último, ¿es la *c* la misma *c* que antes, o la he reemplazado por una idéntica? ¿Qué sucede con la 5? Incluso si simplemente reemplazo la *a* por una *o*, en lugar de reescribir la palabra, hay un instante, justo entre la desaparición de a y la aparición de la *o*, cuando se cierra el hueco y hay (al menos a veces) una onda de realineamiento de la página a medida que la colocación de cada letra sucesiva (incluyendo la 5 que sigue) es recalculada, y luego re-recalculada cuando se inserta la *o*. (Qué barato resulta el cálculo irracional en la modernidad.) En cualquier caso, *todas* las letras que veo ante mí en la pantalla son simples huecos en el trazado de un haz electrónico que explora toda la pantalla sesenta veces por segundo. Si tomo una letra cualquiera y la reemplazo por una idéntica,

[11] Algunos lectores, expertos en tales materias, podrían poner reparos frente a alguna diferencia de signo. Pero incluso esa (defendible) diferencia desaparece si al hacer el intercambio rotamos uno de los electrones 360° (Véase capítulo VI para la explicación).

¿la situación es la *misma* después del reemplazamiento o es simplemente indistinguible de ella? Considerar el segundo punto de vista (es decir, "simplemente indistinguible") como diferente del primero (es decir, "la misma") parece una pérdida de tiempo. Por lo menos parece razonable decir que la situación es la misma cuando las letras son las mismas. Y así sucede con la mecánica cuántica de partículas idénticas. Reemplazar una partícula por otra idéntica es no haber cambiado para nada el estado. La situación debe ser considerada la *misma* que antes. (Sin embargo, como veremos en el capítulo VI, la diferencia *no* es nada trivial en un contexto mecánico-cuántico.)

Los comentarios anteriores sobre la continua renovación de los átomos del cuerpo de una persona se hicieron en el contexto de la física clásica más que en el de la cuántica. Se expresaron como si mantener la individualidad de cada átomo tuviera algún significado. De hecho la física clásica es adecuada y no andamos totalmente errados, a este nivel de descripción, al considerar los átomos como objetos individuales. Siempre que al moverse estén razonablemente bien separados de sus homólogos idénticos, *podremos* referirnos coherentemente a ellos como si mantuvieran sus identidades individuales, puesto que cada átomo puede ser rastreado continuamente de modo que podríamos pensar que no perdemos de vista a ninguno de ellos. Desde el punto de vista de la mecánica cuántica sería sólo una conveniencia de lenguaje el referirnos a la individualidad de los átomos, pero en este nivel nuestra descripción es bastante consistente.

Aceptemos que la individualidad de una persona no tiene nada que ver con la individualidad que pudiéramos atribuir a sus constituyentes biológicos. Más bien está relacionada con la *configuración*, en cierto sentido, de dichos constituyentes, digamos la configuración espacial o espacio-temporal. (Más tarde diremos más sobre esto.) Pero los defensores de la IA fuerte van mucho más lejos. Ellos dirán que si la información contenida en tal configuración puede ser transferida a otra forma desde la cual puede ser recuperada, entonces la individualidad de la persona debe permanecer intacta. Es como las series de letras que acabo de teclear y veo ahora mostradas en la pantalla de mi procesador de textos. Si las borro de la pantalla todavía permanecen codificadas en forma de ciertos minúsculos desplazamientos de carga eléctrica, en una configuración que no tiene una forma geométrica parecida a la de las letras que acabo de teclear. Pero puedo devolverlas en cualquier momento a la pantalla y allí están, como si no hubiese tenido lugar transformación alguna. Si decido guardar lo que acabo de escribir, puedo entonces transferir la información de las secuencias de letras a configuraciones magnéticas en un disco que puedo extraer. Al desconectar la máquina neutralizaré todos los minúsculos (pero significativos) desplazamientos de carga en ella. Mañana podré reinsertar el disco, restablecer los pequeños desplazamientos de carga y mostrar de nuevo la secuencia de letras en la pantalla, como si nada hubiera pasado. Para los defensores de la IA fuerte es "evidente" que la individualidad de una persona puede ser tratada del mismo modo. Dirán que, al igual que con la secuencia de letras en mi pantalla, no se pierde nada de la individualidad de una persona —de hecho, no le habría sucedido nada en absoluto— si su forma física es transferida a algo bastante diferente, por ejemplo, en campos magnéticos sobre un bloque de hierro. Incluso parecen querer decir que la conciencia de una persona persistiría aunque su "información" esté en esta otra forma. Desde este punto de vista, una "conciencia humana" debe considerarse, en realidad, como un elemento de *software*, *y* su manifestación particular como ser humano material debe considerarse como la ejecución de ese *software* por el *hardware* de su cerebro y su cuerpo.

Parece ser que la razón para estas afirmaciones es que, cualquiera que sea la forma material que tome el *hardware* —por ejemplo, algún dispositivo electrónico—, se "plantearía" siempre

cuestiones de *software* (a la manera de la prueba de Turing) y estas respuestas serían idénticas a las que hubiera dado la persona en su estado normal, si suponemos que el *hardware* funciona satisfactoriamente al computar las respuestas a estas cuestiones. ("¿Cómo te sientes esta mañana?" "Oh, muy bien, gracias, aunque tengo un ligero dolor de cabeza". "No sientes, entonces, que hay...ejem...algo extraño en tu identidad personal...o algo parecido?" "No; ¿por qué dices eso? Me parece una pregunta bastante extraña para hacer?" "¿Entonces te sientes la misma persona que eras ayer?" "¡Por supuesto que sí!")

Una idea frecuentemente discutida en este tipo de contexto es la *máquina de teleportación* de la ciencia ficción.[12] Se propone como un medio de "transporte" de, pongamos por caso, un planeta a otro, pero el objeto de la discusión es si realmente existirá tal máquina. En lugar de ser transportado físicamente por una nave espacial en la forma "normal", el hipotético viajero es explorado de arriba a abajo, registrando con todo detalle la localización exacta y la especificación completa de cada átomo y cada electrón de su cuerpo. Toda esta información es entonces emitida (a la velocidad de la luz), mediante una señal electromagnética, al lejano planeta de destino previsto. Allí, la información es recogida o utilizada como instrucciones para ensamblar un duplicado exacto del viajero junto con sus recuerdos, sus intenciones, sus esperanzas y sus sentimientos mas profundos. Al menos eso es lo que se espera, pues cada detalle del estado de su cerebro ha sido fielmente registrado, transmitido y reconstruido Suponiendo que el mecanismo ha funcionado, la copia original del viajero puede ser destruida "sin peligro". Por supuesto, la pregunta es- ¿Es éste *realmente* un método de viajar de un lugar a otro o es simplemente la construcción de un duplicado y el asesinato del original? ¿Estaría *usted* dispuesto a utilizar este método de "viaje", suponiendo que el método se ha mostrado completamente fiable, dentro de sus límites previstos? Si la teleportación *no* es viajar, entonces ¿cuál es la diferencia *de principio* entre esto y caminar simplemente de una habitación a otra? En este último caso ¿no están los átomos de uno en un momento dado proporcionando simplemente la información para las localizaciones de los átomos en el instante siguiente? Después de todo, hemos visto que no hay significado en preservar la identidad de cualquier átomo particular. Ni siquiera tiene significado la cuestión de la identidad de cualquier átomo particular. ¿No constituye cualquier estructura de átomos en movimiento simplemente un tipo de onda de información que se propaga de un lugar a otro? ¿Dónde está la diferencia esencial entre la propagación de la onda que describe a nuestro viajero caminando normalmente de una habitación a otra y la que tiene lugar en el dispositivo teleportador?

Supongamos que es cierto que la teleportación "funciona" realmente, en el sentido de que la propia "conciencia" del viajero es reanimada en la copia de sí mismo en el planeta lejano (suponiendo que esta cuestión tenga un verdadero significado), ¿qué sucedería si la copia *original* del viajero no fuera destruida, como requieren las reglas del juego? ¿Estaría su "conciencia" en dos lugares a la vez? (Trate de imaginar su respuesta cuando le dicen lo siguiente: "¡Oh Dios mío!, ¿de modo que el efecto de la droga que le suministramos antes de colocarle en el teleportador ha desaparecido prematuramente? Esto es un poco desafortunado, pero no importa. De todos modos le gustará saber que el otro usted —ejem, quiero decir el usted *real*, esto es— ha llegado a salvo a Venus, de modo que podemos, ejem, disponer de usted — ejem, quiero decir de la copia *redundante* que hay aquí. Será, por supuesto, totalmente indoloro.") La situación tiene un aire de paradoja.

[12] Véase la Introducción a Hofstadter y Dennett (1981).

¿Hay algo en las leyes de la física que imposibilite *en principio* la teleportación? Quizá, en efecto, no haya nada en principio contra el hecho de transmitir una persona con todo y conciencia por tales medios; en tal caso, el proceso de copiado ¿destruiría inevitablemente el original? Entonces, ¿lo que es imposible en principio es conservar *dos* copias viables?. Creo que, a pesar de lo extravagante de estas consideraciones, hay algo significativo que se puede extraer respecto a la naturaleza física dela conciencia y de la individualidad humanas. Creo que proporcionan un indicador sobre la importancia de la *mecánica cuántica* en la comprensión de los fenómenos mentales. Pero me estoy adelantando. Será necesario volver a estas cuestiones después de que hayamos examinado la estructura de la teoría cuántica en el capítulo VI.

Veamos ahora qué relación guarda el punto de vista de la IA fuerte con la cuestión de la teleportación. Supongamos que en algún lugar entre los dos planetas hay una estación repetidora en la que se almacena temporalmente la información antes de ser retransmitida a su destino final. Por conveniencia, esta información no es almacenada en forma humana sino en algún dispositivo magnético o electrónico. ¿Estaría presente la "conciencia" del viajero en este dispositivo? Los defensores de la IA fuerte tendrán que hacernos creer que así debe ser. Después de todo, dicen ellos, cualquier pregunta que decidiésemos plantear al viajero podría ser respondida en principio por el dispositivo, estableciendo "simplemente" una simulación apropiada de la actividad de su cerebro. El dispositivo contendría toda la información necesaria; el resto sería sólo un asunto de computación. Puesto que el dispositivo respondería a todas las preguntas exactamente como si fuera el viajero, entonces (por la prueba de Turing) *sería* el viajero.

Esto nos lleva de nuevo a la concepción de la IA fuerte según la cual el *hardware* real no es importante en los fenómenos mentales. Esta opinión me parece injustificada. Se basa en la presunción de que el cerebro (o la mente) es, en efecto, una computadora digital. Supone que cuando pensamos no está en juego ningún fenómeno físico concreto que requiriera estructuras físicas concretas (biológicas, químicas) como las que tienen realmente los cerebros.

Se alegará sin duda (desde el punto de vista de la IA fuerte) de que la única suposición que se está haciendo es que los efectos de cualquier fenómeno físico concreto pueden siempre ser exactamente *modelados* mediante cálculos digitales. Estoy totalmente seguro de que la mayoría de los físicos aducirían que esta es una suposición natural en el marco de nuestra comprensión actual de la física. En los últimos capítulos presentaré mi propio punto de vista, contrario a éste (también necesitaré preparar el terreno para decir por qué creo que no vale la pena hacer ninguna suposición al respecto). Pero, por el momento, aceptemos el punto de vista (frecuente) de que toda la física importante *puede* ser siempre modelada mediante cálculos digitales. Entonces, la única suposición real (aparte de las cuestiones de tiempo y capacidad de cálculo) es la "operacional" de que si algo *actúa* plenamente como una entidad consciente, entonces se debe también sostener que ese algo "siente" ser esa entidad.

El punto de vista de la IA fuerte sostiene que, al tratarse "sólo" de una cuestión de *hardware*, toda la física implicada en el funcionamiento del cerebro puede ser simulada mediante la introducción del *software* convertidor apropiado. Si aceptamos el punto de vista operacional, la cuestión descansa entonces en la equivalencia de las máquinas universales de Turing y en el hecho de que cualquier algoritmo puede ser ejecutado por tales máquinas, junto con la presunción de que el cerebro actúa de acuerdo con algún tipo de algoritmo. Ha llegado el momento de ser más explícito sobre estos intrigantes e importantes conceptos.

II. ALGORITMOS Y MÁQUINAS DE TURING

FUNDAMENTOS DEL CONCEPTO DE ALGORITMO

¿QUÉ ES EXACTAMENTE UN ALGORITMO, una máquina de Turing o una máquina universal de Turing? ¿Por qué estos conceptos son cruciales para el punto de vista moderno de lo que podría constituir un "dispositivo pensante"? ¿Existe alguna limitación para lo que un algoritmo pueda hacer, en principio? Para tratar adecuadamente estas cuestiones necesitaremos examinar con cierto detalle la idea de algoritmo y la de máquina de Turing. En las discusiones que siguen tendré que utilizar en ocasiones expresiones matemáticas. Sé muy bien que algunos lectores pueden desanimarse y que otros tal vez se asusten. Si usted es uno de ellos le pido indulgencia y le recomiendo seguir el consejo que di en mi "Nota para el lector" en la página 8. Los argumentos que se dan aquí no requieren conocimientos matemáticos por encima del nivel de escuela elemental, pero para seguirlos en detalle será necesaria una reflexión seria. De hecho, la mayor parte de las descripciones son bastante explícitas y se puede llegar a una buena comprensión siguiendo los detalles. Pero también se puede prescindir de los pasos detallados para sacar simplemente la idea general. Si, por el contrario, usted es experto, le pido de nuevo indulgencia. Espero que se tome la molestia de hojear lo que tengo que decir, y quizá una o dos cosas despierten su interés.

La palabra "algoritmo" procede del nombre del matemático persa del siglo IX Abu Ja'far Mohammed ibn Mûsa al-Khowârizm, autor de un interesante texto matemático, escrito alrededor del año 825 d.C., titulado "Kitab al jabr wa'1-muqabala". El que en la actualidad se escriba "algoritmo", en lugar de la forma antigua, y más aproximada, "algorismo", se debe a una asociación con la palabra "aritmética". También es digno de mención que la palabra "álgebra" procede del árabe "al jabr" que aparece en el título de su libro.

Sin embargo, mucho antes de la aparición del libro de Al- Khowârizm ya se conocían ejemplos de algoritmos. Uno de los más familiares, que data de la época griega antigua (c. 300 a.C.), es el procedimiento hoy conocido como *algoritmo de Euclides* para encontrar el máximo común divisor de dos números. Veamos cómo funciona.

Nos ayudará considerar un par concreto de números, por ejemplo 1365 y 3654. El máximo común divisor es el mayor número entero que es divisor exacto de ambos. Para aplicar el algoritmo de Euclides dividimos uno de los dos números por el otro y tomamos el resto: 3654 entre 1365 cabe a 2 y restan 924 (3654 - 2730). Ahora reemplazamos nuestros números originales por el resto, a saber 924, y el divisor de la operación anterior a saber 1365. Repetimos la operación utilizando ahora este nuevo par de números: 1365 dividido entre 924 cabe a 1 y restan 441.

Esto nos da un nuevo par, 441 y 924, y entonces dividimos 924 entre 441 obteniendo el resto 42 (924 - 882), y así sucesivamente hasta llegar a una división exacta. Escrito en orden todo esto tenemos

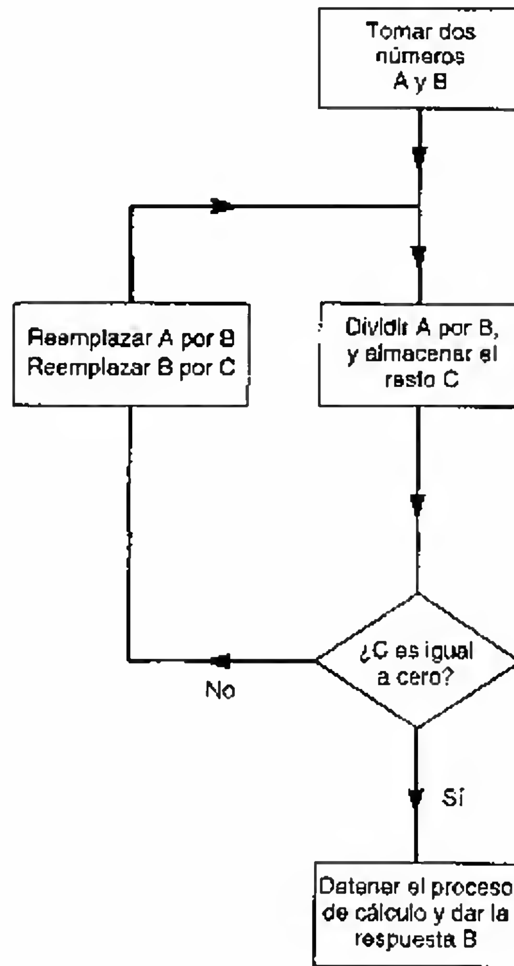
3654 :	1365	da resto 924
1365 :	924	da resto 441
924 :	441	da resto 42
441 :	42	da resto 21
42 :	21	da resto 0

El último número por el que dividimos, es decir 21, es el máximo común divisor buscado.

El algoritmo de Euclides es el *procedimiento sistemático* mediante el cual encontramos este divisor. Hemos aplicado este procedimiento a un par de números particular, pero el procedimiento se aplica a cualquier par de números de cualquier magnitud. Para números muy grandes el procedimiento puede tardar mucho tiempo, y cuanto mayores sean los números más tiempo necesitará. Pero en cualquier caso el procedimiento llegará al final y se obtendrá una respuesta definida en un número finito de pasos. En cada paso está perfectamente claro qué operación debe realizarse, y también es perfectamente clara la decisión de cuándo debe darse por terminado todo el proceso. Además, la descripción del proceso total puede presentarse en términos *finitos*, a pesar de que se aplique a números naturales de tamaño ilimitado. (Los "números naturales" son simplemente los números enteros no negativos[1] 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11...). De hecho, es fácil construir un "organigrama" (finito) para describir la operación lógica completa del algoritmo de Euclides (véase la siguiente figura).

Debe señalarse que este procedimiento no ha sido todavía descompuesto en sus partes más elementales, hemos supuesto implícitamente que ya "sabemos" cómo efectuar la operación básica necesaria para obtener el resto de una división entre dos números naturales A y B arbitrarios. Dicha operación es también algorítmica y es realizada mediante el familiar procedimiento de la división que aprendemos en la escuela.

[1] Estoy siguiendo la terminología moderna usual que ahora incluye el cero entre los "números naturales".



Este procedimiento es en realidad bastante más complicado que el algoritmo de Euclides pero de nuevo podemos construir un organigrama. La mayor complicación reside en que (probablemente) utilizaríamos la notación "decimal" estándar para los números naturales, de modo que necesitaríamos tener una tabla de multiplicar para llevar a cabo estas multiplicaciones. Si hubiéramos utilizado simplemente una serie de *n* marcas de algún tipo para representar el número *n* —por ejemplo, ●●●●● para representar el cinco— entonces encontrar el resto se convierte en una operación algorítmica elemental. Para obtener el resto de A dividido entre B simplemente vamos borrando la serie de marcas que representa B de la serie que representa A, hasta que ya no queden suficientes marcas para realizar de nuevo la operación. La serie de marcas que quede proporciona la respuesta buscada. Por ejemplo, para obtener el resto de la división de diecisiete entre cinco procedemos simplemente a borrar series de ●●●●● de la serie ●●●●●●●●●●●●●●●●● de la forma siguiente

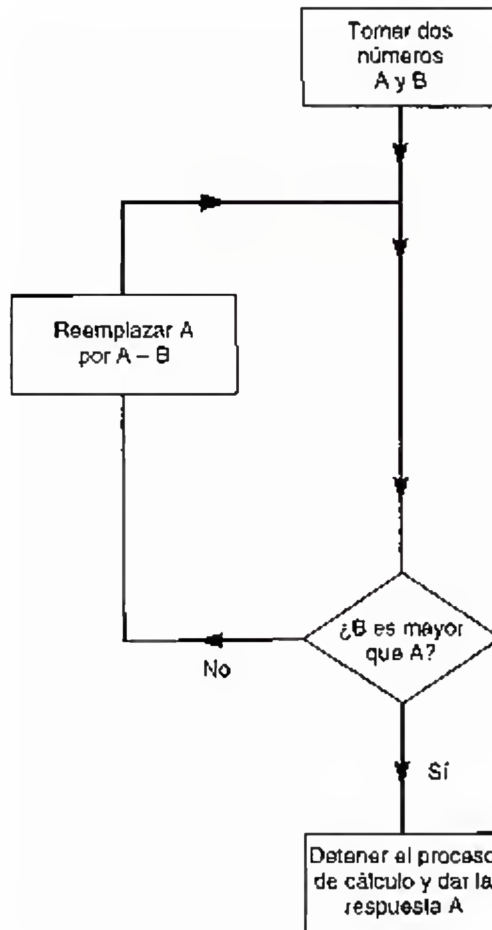
●●●●●●●●●●●●●●●●●

●●●●●●●●●●●●

●●●●●●●

●●

Y evidentemente la respuesta es dos, puesto que ya no podemos realizar la operación otra vez. El organigrama que encuentra el resto de una división, por medio de esta substracción repetida, puede ser el siguiente:



Para completar el organigrama del algoritmo de Euclides, sustituimos el diagrama anterior para encontrar el resto en el cuadro central derecho de nuestro organigrama original. Este tipo de sustitución de un algoritmo dentro de otro es un procedimiento normal en la programación de computadoras. El algoritmo anterior para hallar el resto es un ejemplo de subrutina, es decir, un algoritmo (normalmente conocido con anterioridad) al que se acude y es utilizado por el algoritmo principal como parte de su operación.

Por supuesto, la representación del número N como *n* puntos resulta muy poco eficaz cuando N es muy grande, razón por la cual normalmente utilizamos una notación más compacta como lo es la notación estándar (decimal). Sin embargo, aquí no estamos demasiado interesados en la *eficiencia* de las operaciones o de las notaciones. Más bien estamos interesados en la cuestión de qué operaciones, *en principio*, pueden realizarse algorítmicamente. Lo que es algorítmico lo es en una notación u otra. La única diferencia reside en el detalle y en la complejidad de los dos casos. El algoritmo de Euclides es sólo uno entre los numerosos, a menudo clásicos, procedimientos algorítmicos que florecen en las matemáticas. Pero es de llamar la atención que, a pesar de la antigüedad de muchos algoritmos, la formulación precisa del *concepto general de algoritmo* data sólo de este siglo. De hecho se han dado varias descripciones alternativas de este

concepto, todas ellas en los años treinta. La más directa y convincente, y también la más importante históricamente, es la llamada *máquina de Turing*. Será conveniente para nosotros examinar esta "máquina" con algún detalle. Lo primero que hay que tener en cuenta es que la *máquina de Turing* es un elemento de "matemática abstracta" y no un objeto físico. El concepto fue introducido en 1935 o 1936 por el matemático inglés, extraordinario descifrador de códigos y pionero de las computadoras, Alan Turing (Turing, 1937) para tratar un problema muy general, conocido como el *Entscheidungsproblem*, parcialmente planteado por el gran matemático alemán David Hilbert en 1900, en el Congreso Internacional de Matemáticos en París ("décimo problema de Hilbert") y, de forma más completa, en el Congreso Internacional de Bolonia en 1928. Hilbert buscaba nada menos que un procedimiento algorítmico general para resolver cuestiones matemáticas o mejor dicho, una respuesta a la cuestión de si semejante procedimiento podía o no existir. Hilbert tenía también un programa que pretendía situar las matemáticas sobre una base inatacable, con axiomas y reglas que quedaran establecidas de una vez por todas, pero para cuando Turing produjo su gran obra, dicho programa ya había sufrido un revés decisivo por parte de un sorprendente teorema demostrado en 1931 por el brillante lógico austríaco Kurt Gödel. Consideraremos el teorema de Gödel y su significado en el capítulo IV. El problema de Hilbert que interesaba a Turing (el *Entscheidungsproblem*) iba más allá de cualquier formulación concreta de las matemáticas en términos de sistemas axiomáticos. La pregunta era: ¿existe algún procedimiento mecánico general que pueda, *en principio*, resolver uno s otro todos los problemas de las matemáticas, que pertenezcan a alguna clase bien definida? Parte de la dificultad para resolver esta cuestión insistía en decidir lo que se debe entender por "procedimiento mecánico" El concepto quedaba fuera de las ideas matemáticas comunes de la época. Para poder manejarlo, Turing trató de imaginar cómo podría formalizarse el concepto de "máquina", descomponiendo su modo de operar en términos elementales. Parece claro que también Turing consideraba el cerebro humano como una "máquina" en este sentido, de modo que cualquiera que fuera la actividad que pudiera llevar a cabo un matemático cuando aborda sus problemas, ésta también tendría que entrar en la etiqueta de "procedimientos mecánicos".

Aunque esta visión del pensamiento humano parece haber sido valiosa para Turing a la hora de desarrollar este concepto capital, no estamos obligados ni mucho menos a adherirnos a ella. En realidad, al precisar lo que se debía entender por procedimiento mecánico, Turing mostró que existen algunas operaciones matemáticas perfectamente bien definidas que no pueden ser llamadas mecánicas en ningún sentido. Hay quizá cierta ironía en el hecho de que este aspecto de la obra de Turing puede hoy proporcionarnos indirectamente argumentos contra su propio punto de vista respecto a la naturaleza de los fenómenos mentales. Sin embargo, esto no nos interesa por el momento. Primero necesitamos descubrir cuál es realmente el concepto que tiene Turing de un procedimiento mecánico.

EL CONCEPTO DE TURING

Tratemos de imaginar un dispositivo para llevar a cabo un procedimiento de cálculo (definible en términos finitos). ¿Qué forma tendría tal dispositivo? Debemos permitirnos idealizar un poco y no preocuparnos demasiado por cuestiones prácticas: estamos pensando en una máquina matemáticamente idealizada. Queremos que nuestro dispositivo tenga un conjunto discreto de posibles estados diferentes, en número *finito* (aunque sea un número enorme). Los llamaremos estados *internos* del dispositivo. Sin embargo no queremos limitar el tamaño de los cálculos que

nuestro dispositivo pueda realizar. Recordemos el algoritmo de Euclides descrito anteriormente; no hay límite para la magnitud de los números sobre los que el algoritmo puede actuar. El algoritmo —o el *procedímiento* general de cálculo— es el mismo independientemente de la magnitud de los números. Para números muy grandes el procedimiento puede durar mucho tiempo y necesitar una gran cantidad de papel donde realizar las operaciones, pero el *algoritmo* será el mismo conjunto *finito* de instrucciones.

Así, aunque tenga un número finito de estados internos, nuestro dispositivo debe poder manejar un *input* de cualquier tamaño. Además, el dispositivo dispondrá de un espacio ilimitado de almacenamiento externo ("papel") para sus cálculos, y podrá también producir un *output* de tamaño ilimitado. Puesto que nuestro dispositivo tiene sólo un número finito de estados internos distintos no cabe esperar "*cargar*" todos los datos externos ni todos los resultados de sus propios cálculos. En lugar de ello deberá examinar sólo aquellas partes de los datos o cálculos previos que está manejando *en ese momento*, y realizar con ellas todas las operaciones que sea necesario. Puede anotar, quizá en el espacio de almacenamiento externo, los resultados importantes de esta operación y luego pasar, de una manera predeterminada, a la etapa siguiente de la operación. Es la naturaleza ilimitada del *input*, del espacio de cálculo y del *output* lo que nos dice que estamos considerando solamente una idealización matemática en lugar de algo que pudiera ser realmente construido en la práctica (véase fig. II. 1). Pero es una idealización de gran importancia. Las maravillas de la tecnología moderna de computadoras nos han proporcionado dispositivos de almacenamiento electrónico que pueden considerarse como ilimitados para muchos propósitos prácticos. De hecho, el espacio de almacenamiento que hemos llamado "externo" en la discusión anterior podría ser considerado como una parte real de la estructura interna de una computadora moderna. Es quizá una precisión técnica el que determinada parte del espacio de almacenamiento pueda considerarse interna o externa. *Una* manera de referirse a esta división entre la "parte interna del dispositivo" y la parte "externa" podría ser en términos de *hardware* y *software*. La parte interna sería el *hardware* y la parte externa el *software*. No me voy a atener necesariamente a esta denominación pero, de cualquier forma que se considere, las computadoras electrónicas actuales se aproximan de forma realmente notable a la idealización de Turing.

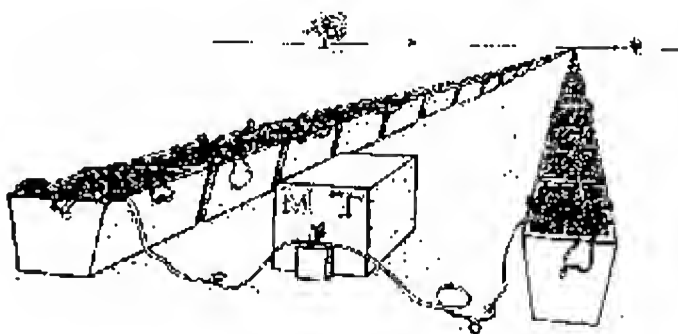


FIGURA II.1. *Una máquina de Turing, en su más estricto sentido, requiere de una cinta infinita*

Turing representaba los datos externos y el espacio de almacenamiento como una cinta sobre la que se hacen marcas. Esta cinta sería utilizada por el dispositivo y leída cuando fuera necesario;

el dispositivo podría, como parte de la operación, mover la cinta hacia adelante o hacia atrás. También podría hacer nuevas marcas en los lugares de la cinta donde fuera necesario y podría borrar las viejas, permitiendo actuar a la *misma* cinta como almacenamiento externo (es decir, como "papel") y como *input*. De hecho resulta conveniente no hacer una distinción clara entre "almacenamiento externo" e *input*, ya que en muchas operaciones los resultados intermedios de un cálculo jugarán el papel de nuevos datos. Recuérdese que en el algoritmo de Euclides reemplazábamos nuestro *input* original (los números A y B) por los resultados de las diferentes etapas de cálculo. Análogamente, la misma cinta puede ser utilizada para el *output* final (es decir, la "respuesta"). La cinta seguirá pasando por el dispositivo hacia adelante y hacia atrás mientras sea necesario hacer nuevos cálculos. Cuando el cálculo haya terminado, el dispositivo se detendrá y la respuesta aparecerá en la parte de la cinta que queda a un lado del dispositivo. Supongamos, para ser concretos, que la respuesta aparece siempre a la izquierda, mientras que los datos numéricos del *input*, junto con los datos del problema a resolver, siempre quedan a la derecha.

Yo personalmente encuentro algo incómodo pensar en nuestro dispositivo finito que mueve hacia adelante y hacia atrás una cinta potencialmente *infinita*. Por ligero que sea el material de que está hecha, una cinta *infinita* será difícil de mover. En su lugar, prefiero pensar la cinta como la representación de un entorno por el cual puede moverse nuestro dispositivo finito. (Por supuesto, con la electrónica moderna ni la cinta ni el dispositivo tienen que moverse realmente en el sentido físico ordinario, pero tal idea de movimiento es una manera conveniente de representar las cosas.) Desde este punto de vista, el dispositivo recibe todo su *input* desde el entorno; utiliza el entorno como el "papel", y al final escribe su *output* en este mismo entorno.

En la imagen de Turing la cinta consiste de una secuencia lineal de cuadros que se considera infinita en ambas direcciones. Cada cuadro de la cinta está en blanco o contiene una sola y única marca.* El uso de cuadros marcados o sin marcar ilustra el hecho de que estamos admitiendo que nuestro entorno (es decir, la cinta) puede ser descompuesto y descrito en términos de elementos *discretos* (y no continuos). Esto es razonable si queremos que nuestro dispositivo funcione de un modo fiable y perfectamente definido. Estamos admitiendo que el entorno sea (potencialmente) infinito como consecuencia de la idealización matemática que estamos utilizando, pero en cualquier caso particular el *input*, el cálculo y el *output* deben ser siempre *finitos*. De este modo, aunque la cinta se considera infinitamente larga, en ella debe haber sólo un número finito de marcas reales. Más allá de un cierto punto en cada dirección la cinta debe estar completamente en blanco.

Indicaremos un cuadro en blanco mediante el símbolo "0" y los marcados mediante el símbolo "1", *v.g.*:

0	0	0	1	1	1	1	0	1	0	0	1	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0

Necesitamos que nuestro dispositivo "lea" la cinta, y supondremos que lo hace cuadro por cuadro, y que después de cada operación se mueve sólo *un* cuadro a la derecha o a la izquierda.

* En realidad, en su descripción original Turing permitía que su cinta estuviera marcada de maneras más complicadas, pero esto no supone ninguna diferencia real. Las marcas complejas podrían descomponerse siempre en series de marcas y espacios en blanco. Me tomaré otras libertades sin importancia con respecto a las especificaciones originales de Turing.

No hay ninguna pérdida de generalidad en esto. Un dispositivo que leyera *n* cuadros cada vez o que se moviera *k* cuadros cada vez, puede modelarse fácilmente mediante otro dispositivo que sólo lea de uno en uno. A partir de *k* movimientos de un cuadro puede construirse un movimiento de *k* cuadros, y almacenando *n* lecturas de un solo cuadro se comportaría como si leyera *n* cuadros a un tiempo.

¿Qué puede hacer, en concreto, un dispositivo semejante? ¿Cuál es la forma más general en la que algo que describiríamos como "mecánico" podría funcionar? Recordemos que los *estados internos* de nuestro dispositivo son finitos en número. Todo lo que necesitamos saber, más allá de esa finitud, es que el comportamiento del dispositivo está completamente determinado por su estado interno y por el *input*. Hemos simplificado este *input* haciendo que sólo sea uno de los dos símbolos "0" o "1". Dado el *input* y su estado inicial, el dispositivo actúa de una forma completamente determinista: cambia de un estado interno a otro (quizá el mismo); reemplaza el 0 o el 1 que acaba de leer por el mismo o por un distinto símbolo 0 o 1; se mueve un cuadro hacia la derecha o la izquierda, finalmente, decide si continuar el cálculo o si terminarlo y detenerse.

Para definir las operaciones de nuestro dispositivo de modo explícito *numeramos* los diferentes estados internos, por ejemplo mediante las etiquetas 0, 1, 2, 3, 4, 5,...; la operación del dispositivo, o máquina de Turing, estará entonces totalmente especificada por una lista de sustituciones, tal como:

00 => 00D
01 => 131I
10 => 651D
11 => 10D
20 => 01D.ALTO
21 => 661I
30 => 370D
.
.
.
2100 => 31I
.
.
.
2581 => 00D.ALTO
2590 => 971D
2591 => 00D.ALTO

La cifra escrita en *grandes* caracteres a la izquierda de la flecha es el símbolo que el dispositivo está leyendo sobre la cinta, y que lo reemplaza por la cifra también en grandes caracteres que aparece al centro del lado derecho. D nos dice que el dispositivo va a moverse un cuadro hacia la *derecha* a lo largo de la cinta, e I nos dice que va a moverse un paso hacia la *izquierda*. (Si, como sucede en la descripción original de Turing, consideramos que lo que se mueve es la cinta en lugar del dispositivo, entonces debemos interpretar D como la instrucción de mover la *cinta* un cuadro hacia la *izquierda*, e I la instrucción de moverla a la *derecha*.) La palabra ALTO indica que el cálculo ha terminado y el dispositivo se detendrá. En concreto, la segunda instrucción 01=>131I nos dice que si el dispositivo está en el estado interno 0 y lee 1 en la cinta, entonces debe cambiar

al estado interno 13, dejar el 1 como un 1 y moverse un cuadro hacia la izquierda a lo largo de la cinta. La última instrucción 2591 =>0OD.ALTO nos dice que si el dispositivo está en el estado 259 y lee 1 en la cinta, entonces debe volver al estado 0, borrar el 1 y escribir un 0 en la cinta, moverse un cuadro hacia la derecha a lo largo de la cinta y dar por terminado el cálculo.

En lugar de utilizar los caracteres numéricos 0, 1, 2, 3, 4, 5,... para etiquetar los estados internos, sería más congruente con la notación para las marcas en la cinta el utilizar símbolos construidos sólo a base de 0 y 1. Podríamos utilizar, si quisiéramos, una serie de *n* símbolos 1 para etiquetar al estado *n*, pero eso es ineficiente. En lugar de ello, utilicemos el sistema *binario* de numeración que actualmente ya resulta familiar:

0=>0,

1=>1,

2=>10,

3=>11,

4=>100,

5=>101,

6=>110,

7=>111,

8=>1000,

9=>1001,

10=>1010,

11=>1011,

12=>1100, etc

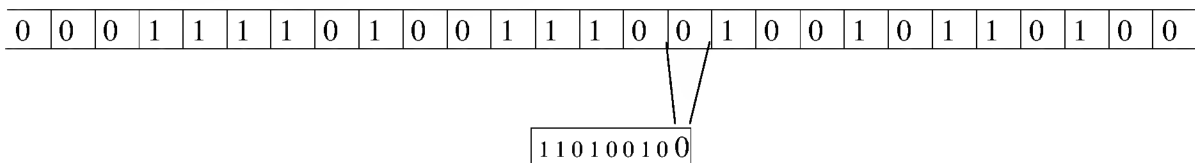
Aquí, el dígito final de la derecha se refiere a las "unidades" igual que en la notación estándar (decimal), pero el dígito inmediatamente anterior se refiere a "doses" en lugar de "decenas". El anterior se refiere a "cuatros" en lugar de "centenas" y el anterior a éste se refiere a "ochos" en lugar de "miles", y así sucesivamente, siendo el valor de cada dígito sucesivo, a medida que nos movemos hacia la izquierda, las sucesivas potencias de dos: 1, 2, 4 (= 2 x 2), 8 (= 2 x 2 x 2), 16 (= 2 x 2 x 2 x 2), 32 (= 2 x 2 x 2 x 2 x 2), etc. (Para otros propósitos, que surgirán posteriormente, nos será útil usar alguna otra base diferente a la del dos o la del diez para representar los números naturales: por ejemplo, en base *tres*, el número decimal 64 se escribiría 2101, pues el valor de cada dígito es ahora una potencia de tres: $64 = (2 \times 3^3) + 3^2 + 1$; *cfr.* capítulo IV, nota a pie de página.) Utilizando la notación binaria para los estados internos, la especificación de la máquina de Turing anterior será ahora:

00 => 00D
01 => 100111
10 => 10000011D
11 => 10D
100 => 01ALTO
101 => 1000010111
110 => 10010100D
. .

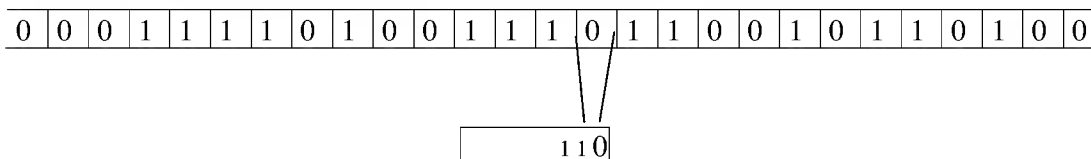
```
        .        .
        .        .
110100100 =>  111I
        .        .
        .        .
        .        .
        .        .
1000000101 =>  00ALTO
1000000110 =>  11000011D
1000000111 =>  00ALTO
```

En el último diagrama he abreviado D.ALTO por ALTO, puesto que podemos suponer que I.ALTO nunca ocurre, de modo que el resultado del último paso de cálculo aparece a la izquierda del dispositivo como parte de la respuesta.

Supongamos que nuestro dispositivo se encuentra en el estado interno representado por la secuencia binaria 11010010 y está a mitad de un cálculo para el que la cinta viene dada como en la p. 43, y aplicamos la instrucción 110100100 =>111I:



El dígito de la cinta que está siendo leído (en este caso el dígito "0") está indicado en caracteres grandes a la derecha de la cadena de símbolos que representan el estado interno. En el caso de una máquina de Turing como la que parcialmente se especifica arriba (y que yo he formado más o menos al azar), el "0" que está siendo leído será reemplazado por un 1 y el estado interno cambiará a "11"; luego el dispositivo se moverá un paso a la izquierda:



El dispositivo está ahora listo para leer otro dígito, de nuevo un "0". De acuerdo con la tabla, ahora deja este "0" inalterado, pero reemplaza el estado interno por "100101" y retrocede a lo largo de la cinta un paso hacia la derecha. Ahora lee "1", y en algún lugar más abajo en la tabla habrá una nueva instrucción sobre cómo debe reemplazar su estado interno, si debe cambiar el dígito que está leyendo, y en qué dirección debe moverse. Continuará así hasta que llegue a un ALTO, punto en el cual (tras moverse un paso más hacia la derecha) imaginamos una campana que suena para avisar al operador que el cálculo ha concluido.

Supondremos que la máquina empieza siempre con estado interno "0" y que toda la cinta que queda a la izquierda del dispositivo de lectura está inicialmente en blanco. Todas las instrucciones y datos se introducen por la derecha. Como ya mencionamos, la información que se introduce toma siempre la forma de una cadena *finita* de 0's y l's, seguidos por cinta en blanco (es decir, todos 0's). Cuando la máquina llega a ALTO, el resultado del cálculo aparece en la cinta a la izquierda del dispositivo de lectura.

Puesto que queremos poder incluir datos numéricos como parte de nuestro *input*, nos gustaría también tener una manera de describir números ordinarios (entiendo aquí por ello los números

naturales 0, 1,2, 3, 4,...) como parte del *input*. Una manera de hacerlo sería simplemente usando una cadena de *n* símbolos l's para representar el número *n* (aunque esto podría plantearnos alguna dificultad con el cero):

1 => 1,2=> 11,3=> 111,4=> 1111,5=> 11111, etc.

Este primitivo sistema de numeración se llama (de forma no muy lógica) sistema *unario*. En este caso el símbolo "0" podría utilizarse como espacio para separar números diferentes. Es importante que dispongamos de este medio de separar números ya que muchos algoritmos actúan sobre *conjuntos* de números y no sobre un solo número. Por ejemplo, para el algoritmo de Euclides nuestro dispositivo necesitará actuar sobre el *par* de números A y B. Se pueden diseñar, sin gran dificultad, máquinas de Turing que ejecuten este algoritmo. Como ejercicio, los lectores aplicados pueden verificar que la siguiente descripción de una máquina de Turing (que llamaré EUC) ejecuta realmente el algoritmo de Euclides cuando se aplica a un par de números unarios separados por un 0:

00 =>00D, 01 => 11I, 10=>101D, 11=> 11I,

100=>10100D, 101 =>110D, 110 =>1000D, 111 =>111 D,

1000 => 1000D, 1001 => 1010D, 1010=>11101 ,

1011=>11111, 1100=>11001, 1101 => 11I,

1110=>11101, 1111=>100011, 10000 => 100101,

10001=>100011, 10010=>100D, 10011 => 11I,

10100=>00ALTO, 10101=>10101D

Antes de embarcarse en tal empresa, sin embargo, sería prudente que el lector empezara por algo mucho más simple, como la máquina de Turing UN + 1:

00 =>00D, 01 => 11D, 10=>01ALTO, 11=> 11D,

que simplemente suma uno a un número unario. Para comprobar que UN + 1 hace precisamente eso, imaginemos que se aplica, por ejemplo, a la cinta

...00000111100000...,

que representa el número 4. Suponemos que el dispositivo está inicialmente en algún lugar a la izquierda de los 1's. Está en el estado interno 0 y lee un 0. De acuerdo con la primera instrucción, deja el 0 como está y se mueve un paso a la derecha quedando en el estado interno 0. Sigue haciendo lo mismo, moviéndose un paso a la derecha, hasta que encuentra el primer 1. Entonces entra en juego la segunda instrucción; deja el 1 como 1 y se mueve otra vez a la derecha, pero ahora en el estado interno 1. De acuerdo con la cuarta instrucción permanece en el estado interno l, dejando el 1 y moviéndose hacia la derecha hasta que llega al primer 0 que sigue a los 1's. La tercera instrucción le dice entonces que cambie ese 0 por un 1, se mueva un paso más a la derecha (recordemos que ALTO representa D.ALTO) y luego se detenga. De este modo, otro 1 se ha añadido a la cadena de l's, y el 4 de nuestro ejemplo se ha cambiado por un 5, como se pretendía.

Como ejercicio algo más complicado se puede comprobar que la máquina UN x 2 definida por:

00 =>00D, 01 => 10D, 10=>1011, 11=> 11D,

100=>110D, 101 =>1000D, 110 =>01ALTO, 111 =>111 D,

1000 => 10111, 1001 => 1011D, 1010=>1011,

1011 =>10111

duplica un número unario.

En el caso de EUC, para tener una idea más clara de lo que supone, Puede ensayarse con algún par explícito de números, digamos 6 y 8. El dispositivo de lectura se encuentra, como antes, en el estado 0 e inicialmente a la izquierda, y la cinta estará ahora marcada inicialmente de la forma:

...0000000000011111101111111100000...

Cuando, después de muchos pasos, la máquina de Turing se detenga, tendremos una cinta marcada:

...000011000000000000...

con el dispositivo de lectura a la derecha de los dígitos distintos de cero. De este modo obtenemos, tal como debía ser, 2 como el máximo común divisor buscado.

La explicación completa de por qué EUC (o, en su caso, UN x 2) hace realmente lo que se supone que hace entraña algunas sutilezas y resultaría más complicado de explicar de lo que es la propia máquina —ciertamente una característica no poco frecuente de los programas de computadora—. (Entender completamente por qué un procedimiento algorítmico hace lo que se le supone, requiere *intuición* y perspicacia. ¿Son algorítmicas las propias intuiciones? Esta pregunta será importante más adelante.) No intentaré dar aquí una explicación de los ejemplos EUC o UN x 2. El lector que los compruebe descubrirá que me he tomado con el auténtico algoritmo de Euclides la pequeña libertad de expresar las cosas de modo más conciso. La descripción de EUC es aún algo complicada pues comprende 22 instrucciones elementales para 11 estados internos distintos. Gran parte de la complicación es de tipo puramente organizativo. Se observará, por ejemplo, que de las 22 instrucciones sólo tres implican en realidad una alteración de las marcas de la cinta. (Incluso para UN x 2 he utilizado 12 instrucciones, la mitad de las cuales implican sólo una alteración de las marcas.)

CODIFICACIÓN BINARIA DE LOS DATOS NUMÉRICOS

El sistema unario es muy poco eficiente para la representación de números de gran tamaño. En consecuencia, utilizaremos normalmente el sistema *binario* de numeración que describimos anteriormente. No obstante, no podemos hacer esto de forma directa, leyendo sencillamente la cinta como un número binario. Tal como están las cosas no habría modo de definir cuándo termina la representación binaria del número y comienza la sucesión infinita de 0's que representa la cinta en blanco hacia la derecha. Necesitamos alguna notación para dar por terminada la descripción binaria de un número. Más aún, a menudo tendremos que introducir *varios* números, como el *par* de números que requiere el algoritmo de Euclides.[2] Tal como está definido no podemos distinguir los *espacios* entre números de los 0's o cadenas de 0's que

[2].Existen muchas otras maneras, bien conocidas para los matemáticos, de codificar pares, tríos, etc., de números como números individuales, aunque son menos convenientes Para nuestro propósito. Por ejemplo, la fórmula 1/2 $((a + b)^2 + 3a + b)$ representa unívocamente los pares de números naturales (a, b) como un simple número natural. ¡Inténtelo!

aparecen como parte de la representación binaria de un solo número. Incluso, tal vez nos gustaría introducir todo tipo de instrucciones complicadas en la cinta *input* además de los números. Para superar estas dificultades adoptaremos un procedimiento que llamaré *contracción*, según el cual cualquier cadena de 0's y l's (con un número total finito de l's) *no* se lee simplemente como un número binario, sino que es reemplazada por una cadena de 0's, l's, 2's, 3's, etc., mediante una instrucción en la que cada dígito de la segunda secuencia es simplemente el número de 1's que hay entre 0's sucesivos de la primera. Así, por ejemplo, la secuencia:

010001011010101101000111010101111100110,

sería reemplazada de la siguiente manera:

```
      010  0  01011010101 1010  0  01110101011111 0  0110
      | | | | |   | | |   | | | |    | | |      | |   |
así:  1 0 0 1 2   1 1 2   1 0 0 3    1 1 4      0 2
```

Podemos ahora leer los números 2, 3, 4,... como marcadores o instrucciones de algún tipo. Así, consideremos el 2 simplemente como una coma que indica el espacio entre dos números, mientras 3, 4, 5,... podrían, a nuestro gusto, representar varias instrucciones o notaciones de interés, tales como "menos", "más", "por", "saltar a la posición del número siguiente", "iterar la operación anterior el siguiente número de veces", etc. Ahora tenemos varias cadenas de 0's y l's que están separadas por dígitos mayores. Las primeras representan números ordinarios escritos en la escala binaria. De este modo, la secuencia anterior se leerá (entendiendo el "2" como "coma")

(número binario 1001) coma (número binario 11) coma...

Utilizando la notación arábiga habitual "9", "3", "4", "0" para los números binarios 1001, 11, 100, 0 respectivamente, obtenemos para toda la secuencia

9, 3, 4 (instrucción 3) 3 (instrucción 4) 0,

Así, este procedimiento nos proporciona un medio para dar por terminada la descripción de un número (y por consiguiente distinguirlo de una extensión infinita de cinta en blanco a la derecha) utilizando simplemente una coma al final. Además, nos permite codificar cualquier secuencia finita de números naturales, escritos en notación binaria, como una *única* secuencia de 0's y l's, en la cual usamos comas para separar los números. Veamos cómo funciona esto en un caso concreto. Consideremos por ejemplo la secuencia

5, 13, 0, 1, 1, 4,

en notación binaria esto es

101, 1101, 0, 1, 1, 100,

que se codifica en la cinta *por expansión* (es decir, el inverso del anterior procedimiento de contracción) como

...000010010110101001011001101011010110100011000...

Para conseguir esta codificación de una forma sencilla y directa podemos hacer sustituciones en nuestra secuencia original de números binarios de la siguiente forma:

0 => 0

1 => 10

, => 110

y luego añadir una cantidad ilimitada de 0's a ambos lados. Se verá más claro cómo se ha aplicado esto a la cinta anterior si la espaciamos

0000 10 0 10 110 10 10 0 10 110 0 110 10 110 10 110 10 0 0 110 00

A esta notación para números y conjuntos de números la llamaremos notación *binaria expandida*. (Por ejemplo, la forma binaria expandida del 13 es 1010010.) Hay que hacer una última puntualización sobre esta codificación. Es sólo una cuestión técnica pero necesaria para la completez.[3] En la representación binaria (o decimal) de los números naturales existe una pequeña redundancia en el hecho de que los 0's colocados a la izquierda de una expresión no "cuentan" — y normalmente se omiten, v.g. 00110010 es el mismo número binario que 110010 (y 0050 es el mismo número decimal que 50). Esta redundancia se extiende al propio número cero que puede escribirse 000, o bien 00 o simplemente 0. De hecho, un espacio en blanco lógicamente también debería denotar cero. En notación ordinaria esto llevaría a gran confusión pero encaja muy bien en la que acabamos de describir. De este modo, un cero entre dos comas puede escribirse también como dos comas juntas („) lo que sería codificado en la cinta como dos pares 11 separados por un solo 0:

...001101100...

Entonces el conjunto anterior de seis números puede escribirse también en notación binaria de la forma

101,1101,,1,1,100,

y ser codificado en la cinta en forma binaria expandida como

...000010010110101001011011010110101101000110OO...

(donde hay un 0 menos que en la secuencia que teníamos antes).

Ahora podemos considerar una máquina de Turing para ejecutar, por ejemplo, el algoritmo de Euclides aplicándolo a pares de números escritos en notación binaria expandida. Por ejemplo, para el par de números 6, 8 que consideramos previamente, en lugar de utilizar

---

[3]. No me he molestado, en lo que precede, en introducir ninguna marca para *iniciar* la secuencia de números (o instrucciones, etc.). Esto no es necesario para el *input*, ya que las cosas empiezan precisamente cuando se encuentra el primer 1. Sin embargo, para el *output* se necesita algo más, puesto que podríamos no saber *a priori* hasta dónde mirar en la cinta *output* para alcanzar el primer 1 (esto es, el que está más a la izquierda). Incluso aunque hubiéramos encontrado una larga cadena de 0's que se extiende hacia la izquierda, esto no garantizaría que no hubiera un 1 aún *más lejos* hacia la izquierda. Podemos adoptar diversos puntos de vista sobre esto. Uno de ellos consistiría en utilizar siempre una marca especial (digamos, codificada por 6 en el procedimiento de contracción) para iniciar el *output* completo. Pero por simplicidad, adoptaré un punto de vista diferente en mis descripciones: el de que siempre se "conoce" cuánta cinta ha sido realmente recorrida por el dispositivo (*v.g.* podemos imaginar que deja una "estela" de algún tipo), de modo que, en principio, no tenemos que examinar una cantidad infinita de cinta para estar seguros de que ha sido revisado todo el *output*.

...000000000001111110111111100000...

como hicimos antes, consideraremos las representaciones binarias de 6 y 8, es decir 110 y 1000 respectivamente. El par es 6,8, esto es, en notación binaria, 110,1000, que, por expansión, se codifica como la cinta

...00000101001101000011000000...

Con este par de números en particular no se gana concisión respecto a la forma unaria. Sin embargo, supongamos que tomamos los números (decimales) 1583169 y 8610. En notación binaria estos serían:

11 00000 101 00001 000001,   10000110100010,

de modo que el par codificado queda como la cinta

...001010000001001000001000000101101000001010010000100110...

que cabe en dos líneas*, mientras que en notación unaria la cinta que representara "1583169, 8610" ocuparía más espacio que todo este libro.

Si quisiéramos, podríamos obtener una máquina de Turing que ejecute el algoritmo de Euclides cuando los números se expresan en notación binaria expandida, añadiendo simplemente a EUC un par apropiado de algoritmos "subrutina" que traduzcan de unario a binario expandido. No obstante, esto no serviría de mucho debido a que la ineficiencia del sistema de numeración unario estaría aún presente "internamente" y quedaría de manifiesto en la lentitud del dispositivo y en la desmedida cantidad de "papel" (a la izquierda de la cinta) que sería necesaria. También puede diseñarse una máquina de Turing más eficiente para el algoritmo de Euclides, que opere íntegramente dentro del binario expandido, pero no sería especialmente ilustrativa para nosotros en este momento.

En lugar de ello, para ilustrar cómo se puede hacer una máquina de Turing que opere con números en expansión binaria, ensayaremos algo mucho más sencillo que el algoritmo de Euclides, a saber: el simple proceso de *sumar uno* a un número natural. Esto puede ser efectuado por la máquina de Turing (que llamaré XN +1):

00 =>00D,   01 => 111,   10=>00D,   11=> 101D,

100=>1101,   101 =>101D,   110 =>01ALTO,   111 =>10001,

1000 => 10111,   1001 => 10011,   1010=>1100D ,

1011=>101D,   1101=>1111D,   1110 => 111D,

1111=>1110D,

Una vez más, el lector aplicado podría dedicarse a comprobar que esta máquina hace realmente lo que debe hacer, aplicándola por ejemplo al número 167, cuya representación binaria es 10100111 y por lo tanto vendría dado por la cinta

* En el original en papel eran efectivamente dos líneas. Al migrarlo al formato electrónico estas se transformaron en solo una. (Nota del revisor)

...000010010001010101011000...

Para sumar uno a un número binario simplemente localizamos el último 0 y lo cambiamos por 1, y luego reemplazamos todos los 1's que siguen por 0's, *v.g.*, 167 + 1 = 168 se escribe en notación binaria

10100111 + 1 = 10101000.

Así, nuestra máquina de Turing de "sumar uno" reemplazaría la cinta citada anterior por

...0000100100100001100000...

que es lo que hace en realidad.

Nótese que incluso la muy elemental operación de sumar uno es un poco complicada con esta notación: utiliza nada menos que quince instrucciones y ocho estados internos diferentes. Por supuesto, las cosas serían mucho más simples en notación unaria puesto que, en tal caso, "sumar uno" significa simplemente extender la cadena de l's con un 1 más, de modo que no es sorprendente que nuestra máquina UN + 1 sea más elemental. Sin embargo, para números muy grandes, UN + 1 sería extraordinariamente lenta debido a la desmedida cantidad de cinta necesaria, y la máquina más complicada XN + 1 , que opera con la notación binaria expandida, más compacta, sería mejor.

Como un inciso, señalo una operación para la cual la máquina de Turing parece realmente más simple en la notación binaria expandida que en la notación unaria, a saber: *multiplicar por dos*. Aquí, la máquina de Turing XN x 2, dada por

00 =>00D, 01 => 10D, 10=>01D, 11=> 100D,

100=>111D, 110 =>11ALTO,

lleva a cabo esta operación en binario expandido, mientras que la correspondiente máquina en notación unaria, UN x 2, que se describió antes, es mucho más complicada.

Todo esto nos da una idea somera de lo que las máquinas de Turing pueden hacer en un nivel muy básico. Como es natural, este dispositivo puede ser muy complicado, sobre todo cuando hay que realizar operaciones de cierta complejidad. ¿Cuál es el alcance final de tales dispositivos? Consideraremos inmediatamente esta cuestión.

LA TESIS DE CHURCH-TURING

Una vez que nos hemos empezado a familiarizar con la construcción de las máquinas de Turing sencillas, resulta fácil convencerse de que las distintas operaciones aritméticas básicas, tales como la suma de dos números, su multiplicación o la elevación de uno a la potencia del otro, pueden ser efectuadas por máquinas de Turing concretas. No sería demasiado complicado explicarlo, pero prefiero no hacerlo en este momento. También pueden darse operaciones cuyo resultado sea un *par* de números naturales, tales como la división con resto u otras en las que el resultado es un conjunto finito pero arbitrariamente grande de números. Por otra parte, pueden construirse máquinas de Turing para las que no se especifique por adelantado qué operación matemática hay que realizar, sino que las instrucciones para ello estén incluidas en la cinta. Pudiéndose dar el caso de que la operación que haya que realizar en una etapa dependa del

resultado de algún cálculo que la máquina tenga que hacer con anterioridad. ("Si la respuesta a ese cálculo es mayor que tal y tal, hacer esto; en caso contrario, hacer lo otro.")

Una vez que nos damos cuenta de que se pueden construir máquinas de Turing que realizan operaciones aritméticas, o simplemente lógicas, es fácil imaginar en qué forma podrían construirse para realizar tareas más complicadas de naturaleza algorítmica. Cuando se ha jugado un rato con tales cosas se convence uno de que puede construirse una máquina de este tipo para realizar *cualquier operación mecánica*. Desde el punto de vista matemático, podemos *definir* una operación mecánica que pueda ser llevada a cabo por una máquina como la de Turing. El sustantivo "algoritmo" y los adjetivos "computable", "recursivo" y "efectivo" son todos ellos usados por los matemáticos para denotar las operaciones mecánicas que pueden ser realizadas por máquinas teóricas de este tipo. Desde el momento que un procedimiento es suficientemente claro y mecánico, resulta razonable creer que se puede encontrar una máquina de Turing que realmente lo realice. Este era, después de todo, el objetivo de nuestra (más bien, la de Turing) discusión introductoria motivada por el concepto de máquina de Turing.

Por otro lado, podría parecer que el diseño de estas máquinas es innecesariamente restrictivo. Imponer que el dispositivo sólo pueda leer cada vez un dígito binario (0 o 1), y moverse solamente un espacio cada vez a lo largo de una *única* cinta unidimensional parece limitante a primera vista. ¿Por qué no permitir cuatro o cinco, o quizá mil cintas separadas, con un gran número de dispositivos de lectura interconectados y funcionando todos a la vez? ¿Por qué no permitir todo un plano de cuadros con 0's v l's (o quizá una disposición tridimensional) en lugar de insistir en una cinta unidimensional? ¿Por qué no permitir símbolos de un alfabeto o de un sistema de numeración más complicado? De hecho, ninguno de estos cambios supone la más mínima ganancia en cuanto a los resultados, aunque alguno de ellos pueda suponer cierta diferencia en cuanto a la economía de las operaciones (como ciertamente sería el caso si permitiéramos más de una cinta). La clase de operaciones realizables, y que por consiguiente caen bajo el rubro de "algoritmos" ("cómputos" o "procedimientos efectivos" u "operaciones recursivas"), sería exactamente la misma que antes aunque ampliáramos la definición de nuestras máquinas con todas estas modificaciones a la vez.

Veamos que no hay *necesidad* de más de una cinta, con tal que el dispositivo siga encontrando en ella tanto espacio como necesite. Para ello tal vez sea necesario cambiar datos de un lugar a otro de la cinta. Esto puede ser "ineficiente" pero no limita las posibilidades del dispositivo.[4] Análogamente, en principio *no* se gana nada (aunque, en ciertas circunstancias, pueda conseguirse una mayor velocidad de cómputo) utilizando más de un dispositivo de Turing en *acción paralela* —una idea que se ha puesto de moda en los últimos años con los intentos de modelar más exactamente el cerebro humano—. Al tener dos dispositivos separados que no se comunican directamente no se consigue más de lo que se obtiene con dos que *sí* se comuniquen; pero si se comunican entonces son, de hecho, un solo dispositivo.

¿Y qué sucede con la restricción de Turing de tener una cinta unidimensional? Si pensamos que esta cinta representa el "entorno", preferiríamos considerarlo, más que como una cinta

[4] Un modo de codificar la información de dos cintas en una sola cinta consiste en intercalar las cintas. Así, las marcas impares en la cinta simple representarían las marcas de la primera cinta, mientras que las marcas pares representarían las marcas de la segunda cinta. Un esquema análogo funciona para tres o más cintas. La "ineficiencia" de este procedimiento se deriva del hecho de que el dispositivo de lectura tendría que estar fintando hacia adelante y atrás y dejando marcadores en la cinta para conservar la huella de dónde está, tanto en las partes pares como en las partes impares de la cinta.

unidimensional, como una superficie plana, o quizá como un espacio tridimensional. Una superficie plana parecería estar más próxima a lo que necesita un "diagrama de flujo" (como en la anterior descripción del algoritmo de Euclides) que una cinta unidimensional. * No obstante, en principio no hay dificultad alguna para escribir la operación de un organigrama en forma "unidimensional" (*v.g.* mediante una descripción verbal del diagrama). La representación plana bidimensional se hace sólo en nombre de nuestra propia conveniencia y facilidad de comprensión y no supone diferencia alguna en los resultados. Siempre es posible codificar el lugar de una marca o un objeto en un plano bidimensional —o incluso en un espacio tridimensional— sobre una cinta unidimensional. (De hecho, utilizar un plano bidimensional es completamente equivalente a utilizar *dos* cintas. Las dos cintas proporcionarían las dos "coordenadas" que serían necesarias para especificar un punto en el plano bidimensional; de modo análogo *tres* cintas pueden actuar como "coordenadas" de un punto en un espacio tridimensional.) Una vez más, esta codificación unidimensional puede ser "ineficiente" pero no limita las posibilidades.

Pese a todo esto podríamos aún preguntarnos si el concepto de máquina de Turing engloba realmente *todas* las operaciones lógicas o matemáticas que llamaríamos "mecánicas". Cuando Turing escribió su artículo original todo esto era mucho menos claro de lo que es hoy, así que creyó necesario presentar sus argumentos con gran detalle. Lo que Turing postulaba rigurosamente encontró apoyo adicional en el hecho de que, en forma independiente (y en realidad un poco antes), el lógico estadounidense Alonzo Church (con la ayuda de S. C. Kleene) había desarrollado un esquema —el cálculo lambda— dirigido también a resolver el *Entscheidungsproblem* de Hilbert. Aunque este esquema mecánico omnicomprensivo no resultaba tan obvio como el de Turing, tenía algunas ventajas en la sorprendente economía de su estructura matemática. (Describiré el notable cálculo de Church al final de este capítulo.) Hubo aún otras propuestas para resolver el problema de Hilbert también independientemente de Turing (véase Gandy, 1988), muy en particular la del lógico polaco-estadounidense Emil Post (un poco después que Turing, pero con ideas mucho más afines a las de Turing que a las de Church). Pronto se demostró que todos estos esquemas eran completamente equivalentes. Esto reforzó considerablemente la idea, que se llegó a conocer como *Tesis de Church-Turing*, de que el concepto de máquina de Turing (o sus equivalentes) definía realmente lo que, en matemáticas, entendemos por procedimiento algorítmico (o efectivo o recursivo o mecánico). Ahora que las computadoras electrónicas de alta velocidad han llegado a ocupar un lugar tan importante en nuestras vidas cotidianas, poca gente parece sentir la necesidad de cuestionar esta tesis en su forma original. En lugar de ello, se ha dirigido la atención al dilema de si los sistemas *físicos* reales (incluyendo los cerebros humanos) —sujetos como están a leyes *físicas* precisas— son capaces de realizar más, menos o exactamente las mismas operaciones lógicas y matemáticas que las máquinas de Turing. Personalmente, acepto de buen grado la forma *matemática* original de la Tesis de Church-Turing. Por otro lado, su relación con el comportamiento de los sistemas físicos reales es un tema aparte que ocupará nuestra atención más adelante en este libro.

* Tal como se han descrito las cosas aquí, este organigrama o "diagrama de flujo" formaría parte del "dispositivo" más que de la "cinta" entorno. Eran los verdaderos números A, B, A - B, etc., los que se representaban en la cinta. Sin embargo, también nos gustaría expresar las características del dispositivo en una forma lineal unidimensional. Como veremos más adelante, cuando tratemos la máquina universal de Turing, existe una relación íntima entre las características de un "dispositivo" particular y la especificación de posibles "datos" (o "programa"). Es por ello conveniente tenerlas ambas en forma unidimensional.

NÚMEROS DIFERENTES DE LOS NATURALES

En la discusión anterior consideramos operaciones con *números naturales, y* señalamos el hecho notable de que máquinas simples de Turing pueden manejar números naturales de tamaño arbitrariamente grande a pesar de que cada máquina tiene un número *finito* fijo de estados internos diferentes. No obstante, a menudo tenemos que trabajar con tipos de números más complicados que aquellos, tales como números negativos, fracciones o números con infinitos decimales. Los números negativos y las fracciones (v.g. números como -597/26) en las que los numeradores y denominadores pueden ser tan grandes como se quiera, pueden ser fácilmente manejados por máquinas de Turing. Todo lo que necesitamos es una codificación apropiada para los signos "-", y "/", y esto puede hacerse utilizando la notación binaria expandida descrita anteriormente (por ejemplo, 3 para "-" y 4 para "/", codificados como 1110 y 11110, respectivamente, en notación binaria expandida). Los números negativos y fracciones se tratan entonces en términos de conjuntos finitos de números naturales, así que no aportan nada nuevo en las cuestiones generales de computabilidad.

Análogamente, las expresiones decimales *finitas* de cualquier longitud tampoco nos aportan nada nuevo, ya que sólo son casos particulares de fracciones. Por ejemplo, la aproximación decimal finita al número irracional *n*, dada por 3.14159265, es simplemente la fracción 314159265/ 100000000. Sin embargo, las expresiones decimales *infinitas*, tales como la expansión *completa*

$$\pi = 3.14159265358979...$$

presentan ciertas dificultades. Ni el *input* ni el *output* de una máquina de Turing pueden ser, estrictamente hablando, números decimales infinitos. Se podría pensar que es posible encontrar una máquina de Turing que produzca uno tras otro, en cadena, *todos* los dígitos sucesivos, 3, 1, 4, 5, 9,... de la expresión de π en el *output* de la cinta, y simplemente debemos permitir que la máquina siga funcionando indefinidamente. Pero esto *no le está permitido* a una máquina de Turing. Debemos esperar a que la máquina se detenga (y suene la campana) antes de poder examinar el *output*. Mientras la máquina no haya alcanzado una instrucción ALTO, el *output* está sujeto a posibles cambios y, por consiguiente, no puede ser dado por válido. Por otro lado, una vez que se ha alcanzado el ALTO, el *output* es necesariamente finito.

Existe, sin embargo, un procedimiento para hacer que una máquina de Turing produzca dígitos uno tras otro de una forma *legítima* muy parecida a ésta. Si queremos generar una expresión decimal infinita, por ejemplo la de *n*, haríamos que una máquina de Turing produjera la parte entera, 3, haciendo que la máquina actúe al nivel 0; luego produciría la primera cifra decimal, 1, haciendo que la máquina actúe al nivel 1; luego la segunda cifra decimal, 4, haciéndola actuar al nivel 2; luego la tercera, 1, haciéndola actuar al nivel 3, y así sucesivamente. De hecho *existe* una máquina de Turing para producir de *esta* manera la expresión decimal completa de π, aunque sería algo complicado definirla explícitamente. Una puntualización similar es aplicable a muchos otros números irracionales, tales como $\sqrt{2}$ = 1.414213562... Sucede sin embargo que, curiosamente, algunos irracionales no pueden ser producidos por ninguna máquina de Turing, como veremos en el próximo capítulo. Los números que *pueden* ser generados de esta forma se llaman *computables* (Turing, 1937). Los que no pueden (en realidad la inmensa mayoría) se llaman *no computables.* Volveré a este tema, y a otros relacionados, en los últimos capítulos. Resultará de importancia en la cuestión de si un *objeto físico real* (*v.g.* un cerebro humano) puede describirse adecuadamente, según nuestras teorías físicas, en términos de estructuras matemáticas computables.

El tema de la computabilidad es de gran importancia en matemáticas. No deberíamos pensarlo como algo que sólo concierne a los *números* como tales. Podemos tener máquinas de Turing que operen directamente sobre *fórmulas matemáticas* como, por ejemplo, expresiones algebraicas o trigonométricas, o que lleven a cabo manipulaciones formales de cálculo. Todo lo que se necesita es una forma precisa de codificación en secuencias de 0's y 1 's de todos los símbolos matemáticos involucrados y, a continuación, ya podemos aplicar el concepto de máquina de Turing. Después de todo, esto es lo que Turing tenía en mente cuando abordó el *Entscheidungsproblem*, que reclamaba un procedimiento algorítmico para responder cuestiones matemáticas de naturaleza *general*. Volveremos a esto en breve.

LA MÁQUINA UNIVERSAL DE TURING

Todavía no he descrito el concepto de máquina *universal* de Turing. No es demasiado difícil enunciar el principio que hay detrás, aunque los detalles son complicados. La idea básica consiste en codificar la lista de instrucciones para una máquina de Turing arbitraria *T* en una cadena de 0's v 1's que pueda ser representada en una cinta. Esta cinta se utiliza a continuación como la parte inicial del *input* de alguna máquina de Turing *particular U* —que llamaremos máquina universal de Turing— que actúa sobre el resto del *input* de la misma forma que lo hubiera hecho *T*. La máquina universal de Turing es un imitador universal. La parte inicial de la cinta proporciona a la máquina *U* toda la información que necesita para imitar exactamente a cualquier máquina *T*.

Para ver cómo funciona, necesitamos en primer lugar un modo sistemático de *numerar* máquinas de Turing. Consideremos la lista de instrucciones que define a alguna máquina de Turing particular, por ejemplo, una de las descritas más arriba. Debemos codificar esta lista en una cadena de 0's y l's siguiendo un esquema preciso. Esto puede hacerse con la ayuda del procedimiento de "contracción" que adoptamos antes. En efecto, si representamos los símbolos D, I, ALTO, la flecha (=>) y la coma, mediante los números 2, 3, 4, 5 y 6, respectivamente, podemos codificarlos como contracciones por 110, 1110, 11110, 111110 y 1111110. Entonces los dígitos 0 y 1, codificados como 0 y 10, respectivamente pueden ser utilizados en las cadenas reales de estos símbolos que aparecen en la tabla. No necesitamos una notación diferente para distinguir en la tabla de la máquina de Turing las cifras en caracteres grandes 0 y 1 de las más pequeñas en negrita, ya que la posición de los dígitos grandes en el extremo de la numeración binaria es suficiente para distinguirlos de los demás. Así, por ejemplo, 1101 se leería como el número binario 1101 y se codificaría en la cinta como 1010010. En particular, 00 se leería como 00, que puede codificarse sin ambigüedad como 0, o como un símbolo que no se haya usado. Podemos economizar mucho trabajo si no codificamos ninguna flecha ni ninguno de los símbolos que les preceden, basándonos, en cambio, en el ordenamiento numérico de las instrucciones para especificar cuáles deben ser estos símbolos, aunque para este procedimiento debemos asegurarnos de que no haya huecos en el ordenamiento, añadiendo algunas órdenes "mudas" donde sea necesario. (Por ejemplo, la máquina de Turing XN + 1 no tiene ninguna orden que nos diga qué hacer con 1100 ya que esta combinación no aparece nunca en el funcionamiento de la máquina, de modo que debemos insertar una orden "muda", por ejemplo 1100 => 00D, que Puede incorporarse a la lista sin cambiar nada. Análogamente deberíamos insertar 101 => 00D en la máquina XN x 2.). Sin estas órdenes "mudas", la codificación de las instrucciones subsiguientes sería equivocada. En realidad no necesitamos la coma al final de

cada instrucción, como aparece, puesto que los símbolos I o D bastan para separar las instrucciones. En consecuencia, adoptamos sencillamente la siguiente codificación:

0 para 0 o 0, 10 para 1 o 1, 110 para D, 1110 para I,

11110 para ALTO

Como ejemplo, vamos a codificar la máquina de Turing XN + 1 (con la instrucción 1100=>00D insertada). Prescindiendo de las flechas, de los dígitos que los preceden, y de las comas, tenemos:

00D 11D 00D 101D 110I 101D 01ALTO 1000I 1011I

1001I 1100D 101D 00D 1111D 111D 1110D.

Podemos mejorar el procedimiento prescindiendo de todos los 00 y reemplazando cada 01 por un simple 1, según lo que hemos dicho antes, para obtener

D11DD101D110I101D1ALTO1000I1011I1001I1100D101DD1111D111D1110D.

Esto se codifica como la secuencia en la cinta

11010101101101001011010100111010010110101111010000111010010101110100010111010 100011010010110110101010101101010101101010100110

Para simplificar un poco más, podemos también borrar siempre el 110 inicial (junto con la cinta infinita en blanco que le precede) puesto que esto significa 00D, que representa la instrucción inicial 00 => 00D que he supuesto implícitamente común a *todas* las máquinas de Turing —de modo que el dispositivo puede empezar a funcionar arbitrariamente lejos hacia la izquierda de las marcas sobre la cinta e ir hacia la derecha hasta llegar a la primera marca— y siempre podemos borrar el 110 final (y la implícita secuencia infinita de 0's que se supone que le sigue) ya que todas las descripciones de máquinas de Turing deben acabar de esta forma (pues todas terminan con D, I o ALTO). El *número binario* resultante es el *número* de la máquina de Turing, que en el caso de XN + 1 es:

10101101101001011010100111010010110101111010000111010010101110100010111010100 011010010110110101010101101010101101010100

En notación decimal estándar este número en particular es

450 813 704 461 563 958 982 113 775 643 437 908.

A veces nos referimos, de forma un tanto imprecisa, a la máquina de Turing cuyo número es *n*, como la n-ésima máquina de Turing, denotada por T_n..

Así, XN + 1 es la 450 813 704 461 563 958 982 113 775 643 437 908-ésima máquina de Turing.

Parece un hecho sorprendente que tengamos que ir tan lejos en la "lista" de máquinas de Turing antes de encontrar la que realiza una operación tan trivial como la de sumar uno (en la notación binaria expandida) a un número natural. (No creo haber sido especialmente torpe en mi codificación, aunque sí veo posibles mejoras menores.) En realidad existen algunas máquinas de Turing interesantes con números menores. Por ejemplo, UN + 1 tiene el número binario

101011010111101010

que simplemente es 177 642 en notación decimal. En consecuencia, la muy trivial máquina de Turing UN + 1, que coloca un 1 adicional al final de una secuencia de 1's, es la 177 642-ésima máquina de Turing. A modo de curiosidad podemos señalar que multiplicar por dos queda en alguna parte entre estas dos en la lista de máquinas de Turing, en cualquiera de las notaciones, pues encontramos que el número de XN x 2 es 10 389 728 107 mientras que el de UN x 2 es 1 492 923 420 919 872 026 917 547 669.

Quizá no le sorprenda saber, en vista de la magnitud de estos números, que la inmensa mayoría de los números naturales no dan máquinas de Turing que trabajen en forma alguna. Hagamos la lista de las trece primeras máquinas de Turing, según esta numeración:

T_0:	00	=>	00D,	01	=>	00D,	
T_1:	00	=>	00D,	01	=>	00I,	
T_2:	00	=>	00D,	01	=>	01D,	
T_3:	00	=>	00D,	01	=>	00ALTO,	
T_4:	00	=>	00D,	01	=>	10D,	
T_5:	00	=>	00D,	01	=>	01I,	
T_6:	00	=>	00D,	01	=>	00D,	10=>00D
T_7:	00	=>	00D,	01	=>	???,	
T_8:	00	=>	00D,	01	=>	100D,	
T_9:	00	=>	00D,	01	=>	10I,	
T_{10}:	00	=>	00D,	01	=>	11D,	
T_{11}:	00	=>	00D,	01	=>	01ALTO,	
T_{12}:	00	=>	00D,	01	=>	00D,	10=>00D

De éstas, T_0 simplemente se mueve hacia la derecha borrando todo lo que encuentra, sin detenerse nunca ni volver atrás. La máquina T_1 consigue finalmente el mismo efecto pero de una manera más torpe, saltando hacia atrás cada vez que borra una marca de la cinta. La máquina T_2, al igual que la T_0, también se mueve incesantemente hacia la derecha, pero es más respetuosa y deja todo tal como estaba. Ninguna de las dos sirve como máquina de Turing ya que no se detienen nunca. T_3 es la primera máquina respetable: se detiene, modestamente, después de cambiar el primer 1 (el más a la izquierda) por un 0.

T_4 tropieza con un serio problema. Una vez que encuentra su primer 1 en la cinta entra en un estado interno para el que no hay listado, de modo que no tiene instrucciones sobre lo que debe hacer a continuación. T_8, T_9 *y* T_{10} tropiezan con el mismo problema. La dificultad con T_7 es aún más grave. La cadena de 0's y l's que la codifica incluye una secuencia de *cinco* l's sucesivos: 110111110. No existe interpretación para semejante secuencia, así que se quedará bloqueada en cuanto encuentre su primer 1. (Llamaré a T_7 a cualquier otra máquina T_n para la que la expresión binaria de *n* contenga una secuencia de más de cuatro l's, una máquina *no especificada correctamente.*) Las máquinas T_5, T_6 y T_{12} tropiezan con problemas similares a los de T_0, T_1 o T_2. Sencillamente continúan indefinidamente sin detenerse nunca. Las máquinas T_0, T_1, T_2, T_4, T_5, T_6, T_7, T_8, T_9, T_{10} y T_{12} son inútiles. Sólo T_3 *y* T_{11} son máquinas de Turing que funcionan, y no muy interesantes por cierto. T_{11} es aún más modesta que T_3: se detiene al primer encuentro con un 1 y no cambia nada.

Señalemos que hay también una redundancia en nuestra lista. La máquina T_{12} es idéntica a T_6, *y* también idéntica en actuación a T_0, puesto que el estado interno 1 de T_6 y T_{12} nunca interviene. No tenemos que preocuparnos por esta redundancia ni por la proliferación de máquinas de Turing inútiles en la lista. Sería desde luego posible mejorar nuestra codificación para que se

eliminaran muchas de las máquinas inútiles y se redujera considerablemente la redundancia. Todo esto se haría a expensas de la simplicidad de nuestra pobre máquina universal de Turing que tiene que descifrar el código y tratar de ser la máquina de Turing T_n cuyo número n está leyendo. Tal vez merecería la pena hacerlo si pudiéramos eliminar *todas* las máquinas inútiles (o todas las redundancias). Pero, como veremos en un momento, esto *no* es posible. Por consiguiente, dejemos nuestra codificación como está. Será conveniente interpretar una cinta con su serie de marcas, *v.g.*

...0001101110010000...

como la representación binaria de algún número. Recordemos que los 0's continúan indefinidamente en ambas direcciones, pero que hay sólo un número finito de 1's. Estoy suponiendo también que el número de 1's es distinto de cero (esto es, hay al menos un 1). Podríamos decidir leer la cadena finita de símbolos entre el primer y el último 1 (inclusive), que en el caso anterior es

110111001,

como la notación binaria de un número natural (aquí 441, en notación decimal). Sin embargo, este procedimiento sólo nos daría números *impares* (números cuya notación binaria termina con un 1) y pretendemos representar *todos* los números naturales. En consecuencia, adoptamos el sencillo expediente de eliminar el 1 final (que se toma como un simple marcador que señala la terminación de la expresión) y leer lo que queda como un número binario.[5] Así, para el ejemplo anterior, tenemos el número binario

11011100,

que en notación decimal es 220. Este método tiene la ventaja de que el cero también está representado como una cinta marcada, a saber

...0000001000000...

Consideremos la acción de la máquina de Turing T_n sobre alguna cadena (finita) de 0's y 1's en una cinta que introducimos por la derecha. Será conveniente considerar también esta cinta como la representación binaria de algún número, digamos m, según el esquema dado más arriba. Supongamos que tras una serie de pasos la máquina T_n finalmente se detiene (es decir, llega a ALTO). La cadena de dígitos binarios que la máquina ha producido a la izquierda es la respuesta al cálculo. La leemos también como la representación binaria de un número, digamos p. Escribiremos esta relación, que expresa el hecho de que cuando la n-ésima máquina de Turing actúa sobre m produce p, como:

$$T_n(m) = p.$$

Consideremos ahora esta relación en una forma ligeramente diferente. Imaginemos que expresa una operación particular que se aplica al *par* de números n *y* m para dar lugar al número p. (Así, dados los *dos* números n y m podemos calcular a partir de ellos p, viendo lo que la n-ésima máquina de Turing hace con m.) Esta operación particular es un procedimiento totalmente algorítmico y, por lo tanto, puede ser llevado a cabo por *una* máquina de Turing U; esto es, U

[5] Este procedimiento se refiere sólo al modo en que una cinta marcada puede interpretarse como un número natural. No altera los números de nuestras máquinas de Turing específicas, tales como EUC o XN + 1

actúa sobre el *par* (n,m) para producir p. Puesto que la máquina de Turing tiene que actuar sobre *ambos*, n *y* m, para producir el resultado simple p, necesitaremos algún modo de codificar el par (n, m) en la cinta *única*. Para esto podemos suponer que n está escrito en la notación binaria ordinaria y que termina con la secuencia 111110. (Recordemos que el número binario de toda máquina de Turing correctamente especificada es una secuencia formada sólo a base de 0's, 10's, 110's, 1110's y 11110's, y por consiguiente no contiene ninguna secuencia de más de cuatro 1's. De modo que si T_n es una máquina correctamente especificada, la aparición de 111110 significa verdaderamente que la descripción del número n ha terminado.) Todo lo que hay a continuación es simplemente la cinta, representada por m de acuerdo con nuestra prescripción anterior (es decir, el número binario m seguido inmediatamente de 1000...). Entonces esta segunda parte es sencillamente la cinta sobre la que se supone que actúa T_n. *A* modo de ejemplo, si tomamos n = 11 y m = 6 tenemos que la cinta sobre la que U tiene que actuar posee la secuencia de marcas

...0001011111110110100000...

Ésta está formada por:

...0000	(cinta en blanco inicial)
1011	(representación binaria de 11)
111110	(termina n)
110	(representación binaria de 6)
10000...	(resto de la cinta)

Lo que tendría que hacer la máquina *de* Turing U, en cada paso de la operación de T_n sobre m, sería examinar la estructura de la serie de dígitos en la expresión de n de modo que pueda hacerse la sustitución apropiada de los dígitos de m (esto es, la "cinta" de T_n). De hecho, no es difícil (aunque sí tedioso) ver cómo se podría construir realmente una máquina semejante. Su propia lista de instrucciones nos estaría proporcionando un medio de leer la entrada apropiada en dicha "lista", entrada que está codificada en el número n, en cada paso de la aplicación a los dígitos de la "cinta", tal como figuran en m. Por supuesto que habría una serie de idas y vueltas, atrás y adelante de los dígitos de m a los de n y viceversa, y el procedimiento se haría desesperadamente lento. De todas formas, puede darse sin duda una lista de instrucciones para una máquina semejante; y llamamos a esa máquina una máquina *universal* de Turing. Al indicar la acción de esta máquina sobre el par de números n *y* m por $U(n,m)$, tenemos:

$$U(n,m)= T_n(m)$$

para cada (n,m) para el que T_n es una máquina de Turing correctamente especificada.[6] La máquina U, cuando se alimenta inicialmente con el número n, imita exactamente a la n-ésima máquina de Turing.

Puesto que U es una máquina de Turing, tendrá también un número; es decir, tendremos

$$U=T_u,$$

para algún número u. ¿Qué tan grande es u? De hecho podemos tomar *exactamente*

[6] Si T_n *no* está correctamente especificada, entonces U procederá como si el número Para n hubiera terminado en cuanto se alcanza la primera cadena de cuatro 1s o más en la expresión binaria de n. Leerá el resto de esta expresión como parte de la cinta para m, de modo que procederá a ejecutar un cálculo sin sentido. Podríamos eliminar esta característica, si quisiéramos, disponiendo que n se exprese en notación binaria *expandida*. No he querido hacer esto para no complicar más la descripción de la pobre máquina universal de Turing U.

u=7244855335339317577198395039615711237952360672556559631108144796606505059404241090310483613632359365644443458382226883278767626556144692814117715017842551707554085657689753346356942478488597046934725739988582283827795294683460521061169835945938791885546326440925525505820555989451890716537414896033096753020431553625034984529832320651583047664142130708819329717234151056980262734686429921838172157333482823073453713421475059740345184372359593090640024321077342178851492760797597634415123079586396354492269159479654614711345700145048167337562172573464522731054482980784965126988788964569760906634204477989021914437932830019493570963921703904833270882596201301773727202718625919914428275437422351355675134084222299889374410534305471044368695876405178128019437530813870639942772823156425289237514565443899052780793241144826142357286193118332610656122755531810207511085337633806031082361675045635852164214869542347187426437544428790062485827091240422076538754264454133451748566291574299909502623009733738137724162172747723610206786854002893566085696822620141982486216989026091309402985706001743006700868967590344734174127874255812015493663938996905817738591654055356704092821332221631410978710814599786695997045096818419062994436560151454904880922084480034822492077304030431884298993931352668823496621019471619107014619685231928474820344958977095535611070275817487333272966789987984732840981907648512726310017401667873634776058572450369644348979920344899974556624029374876688397514044516657077500605138839916688140725455446652220507242623923792115253181625125363050931728631422004064571305275802307665183351995689139748137504926429605010013651980186945639498

(o alguna otra posibilidad por lo menos de este orden de magnitud). Sin duda, este número parece escalofriante y en efecto *es* escalofriante, pero no veo cómo reducirlo significativamente. Los procedimientos de codificación y las especificaciones que he dado para máquinas de Turing son bastante razonables y simples pero, a pesar de todo, hemos llegado inevitablemente a este enorme número para la codificación de una máquina universal de Turing real.[7]

[7] Debo agradecer a David Deutsch el que haya derivado la forma decimal de la descripción binaria para *u* que he calculado más abajo. También le agradezco el haber comprobado que este valor binario de *u* da efectivamente una máquina universal de Turing. El valor binario para *u* es de hecho:
1000000001011101001101000100101010110100011010001010000011010100110100010101001011010000110100010100101011010010011101001010010010111010100011101010100100101011101010100110100010100010101101000001101001000001010110100010011101001010000101011101001000111010010101000010111010010100110100001000011101010000111010100001001001110100010101011010100101011010000011010101001011010010010001101000000001101000000111010100101010101110100001001110100101010101010101110100001010101110100001010001011101000101001101001000010100110100101001001101001000101101010001011101001001010111010010100011101010010100100111010101010000110100101010101110101001000101101010000101101010001001101010101010001011010010101001001011010100100101110101010010101110101001010011010101000011101000100100101011101010100101011101010100000111010100100000110101010100101110101001010110100010010001110100000000111010010100101010101110100101001001010111010000001010111010000100011101000000101010011101000010100111010000001000101110100010000111010000010010100111010001000010110100010100101110100010100101101001000000101101000101010010011010001010101011101001000000111010010010101010111

He dicho que todas las computadoras modernas de tipo general son en efecto, máquinas universales de Turing. No quiero decir que el diseño lógico de estos ordenadores necesite parecerse mucho a la descripción de una máquina universal de Turing que acabo de dar. Lo importante es simplemente que, si antes que nada suministramos a cualquier máquina universal de Turing un programa apropiado (la parte inicial de la cinta *input*), podemos hacer que imite el comportamiento de cualquier máquina de Turing. En la descripción anterior, el programa toma la forma

10101001101010100101001011010101001101001001010111010011010010000010110100010
10101000111010010000101011010000001001101001000100101110100100001101010000010
01011101001001010011010010010101011010011010010010100101101001101001010000010
11010010000011101010010011010101010000101110100101000010111010010101010111010
10001001011010010011101001010100010111010001001110101000010110100100111010010
10101010111010010001110100101010100101110100100011101010000010101011100110101
00000101101001001110101000000101110100101101010000010101101001010010111010100
00100101110100001101010001000010110101001101010001000101101010101001011101010
00101001011010001010101011101001000010101101010001011101010010010101011101010
10010010111010100011101010001110101001001001011101010001110101001010001011101
01000101110101000010010111010100011101000101000101110100101001011101010010101
00101110100101010101010110101000010101010110100001001110100001010101010111010
10100010101110101010001010111010000001110101010001001011101000000111010101001
00010111010100000011010100001011010000001110100100000010111010100011101010010
001010111010100110101010100010101101000001101010101001010101101000000010011010
10101001001110101001101010101001001011010100110100100100111010000011010101010
10010101101010001001101000101001010101110100000110101010101010010110100010001
11010001010101010101101000100011101000010101110100010010000111010011010000000
10011101000000100101110100010001010011101000000100101110100101010101001011010
00010101010111010001001010010111010000010001011101010100101101000100010011101
00000100101011101000000101010110100001000111001111010000100000111010000100100
11101000001010010111010000010100101101000010001010111010000100010011010001000

10101010100110100100010101101001001001011010000000010110100000100011010000010010110100000000011010010100
0101110100101010001101001010010101101000001001110100101010010110100100111010100000010101110101000000110
1010100010101011010010101011010100001010111010100100101011101010001001011010100100001011101000000111010
1001000101101010010100110101010001011101010010100101110101010000010111010101000001011101000000111010101
0000101011101001010101101010100001011101010001010101110101010010010111010101010000111010100000001110100
1001000011010010010001011010101010100111010000000010110100100001101010101010010111010010000110100100010
10101110100001000111010001000011101000011010000000010110100000100101110101010010101011010001000100101110
1000001001110101010011010000010101011010000100001110100100001000111010101010101001110100001001001110100
0100100001110100001010010110100001010000111010101010101011101000100100110100010010011010100101001011101
00010001010111010000000011101000100100101110100110100100100001011010101010011010001010001011101000011010
100001000101

El lector animoso, que tenga una computadora en casa, puede ocuparse en comprobar, utilizando las recetas dadas en el texto, que el código anterior da realmente una acción de una máquina universal de Turing aplicándola a diversas máquinas de Turing de números sencillos.

Hubiera sido posible una rebaja importante del valor de u con una especificación diferente para una máquina de Turing. Por ejemplo, podríamos prescindir de ALTO y adoptar en su lugar la regla de que la máquina se para cuando vuelve a entrar en el estado interno o después de que haya estado en algún otro estado interno. Esto no sería una gran ganancia (suponiendo que lo fuese). Resultaría una ganancia mayor si permitiésemos cintas con otras marcas además de 0 o 1. Efectivamente se han descrito en la literatura máquinas de Turing de apariencia muy concisa, pero esta concisión es algo decepcionante ya que dependen de codificaciones extraordinariamente complejas para las descripciones de las máquinas de Turing en general.

01110101111010000100100101110100001001001011101000000001010111010000101010001 1
010001001011101000010000011101000010011101000100000101110101010010110100010 00
001011101000010101010111010000001010101110100010000101011101000100001010 11101
0010000001110101001001001101000000101011101000100010010111010101000011 10101001
01011010010101010000110100000101001101000000001110100000100100111010010 1101001
000101001011010101001101000101001001011010101001101000101010001011001 10101001
0010111010101001101 0001010101010110011010100010101011001101001 0001010

de un simple número (el número n), pero hay otros procedimientos posibles y muchas variaciones sobre la forma original de Turing. De hecho, en mi propia descripción me he desviado algo de la que dio Turing originalmente, sin embargo, ninguna de estas diferencias es importante para nuestras actuales necesidades.

LA INSOLUBILIDAD DEL PROBLEMA DE HILBERT

Llegamos ahora al objetivo para el que Turing desarrolló originalmente sus ideas, la resolución del muy general *Entscheidungsproblem* de Hilbert: ¿existe algún procedimiento mecánico para responder a todas las cuestiones matemáticas dentro de un amplio, pero bien definido marco? Turing descubrió que podía plantear la pregunta en términos de decidir si la n-ésima máquina de Turing se *detendrá* o no cuando actúe sobre el número m. Esto fue llamado el *problema de la detención*. Resulta fácil construir una lista de instrucciones de acuerdo con las cuales la máquina no se detenga para *ningún* número m (por ejemplo, $n = 1$ o 2, como sucede en los casos ya señalados o en cualquier otro en el que no haya ninguna instrucción ALTO). Existen también muchas listas de instrucciones de acuerdo con las cuales la máquina siempre se parará, cualquiera que sea el número dado (*v.g.* $n = 11$); y algunas máquinas que se pararán para unos números pero no para otros. Sería correcto decir que un presunto algoritmo no es de mucha utilidad si sigue actuando indefinidamente sin detenerse nunca. De hecho, ése no sería un algoritmo. Por ello, es una cuestión importante el poder decidir si T_n aplicada a m dará o no una respuesta. Si *no* lo hace (esto es, si el cálculo *no* se acaba), escribiremos

$$T_n(m) = \square.$$

Se incluirán en esta notación aquellas situaciones en las que la máquina de Turing abandona un problema debido a que no encuentra ninguna instrucción que le diga lo que tiene que hacer —como sucede con las máquinas inútiles T_4 *y* T_7, consideradas más arriba—. También, desgraciadamente, nuestra máquina aparentemente adecuada T_3 debe ser ahora considerada inútil: $T_3(m) = \square$., debido a que el resultado de la acción de T_3 es siempre una cinta en blanco, y al menos necesitamos un 1 en el *output* para que sea asignado un número al resultado. La máquina T_{11} en cambio, sí es legítima puesto que produce un 1 (sólo uno). Este *output* es la cinta numerada 0, de modo que tenemos $T_{11}(m) = 0$ para toda m. Ahora bien, sería importante desde el punto de vista matemático, poder decidir cuándo se va a parar una máquina de Turing. Consideremos por ejemplo la ecuación:

$$(x + 1)^{w+3} + (y + 1)^{w+3} = (z + 1)^{w+3}.$$

(Si las ecuaciones matemáticas le molestan, no se vaya a desanimar. Esta ecuación es sólo un ejemplo y no es necesario comprenderla en detalle.) Esta ecuación está relacionada con un famoso problema no resuelto en matemáticas —quizá el más famoso de todos—. El problema es el siguiente: ¿existe un conjunto de números naturales *w*, *x*, *y*, *z* que satisfaga esta ecuación?[*] El famoso enunciado conocido como "último teorema de Fermat", escrito en un margen de la *Aritmética* de Diofanto por el gran matemático francés del siglo XVII Pierre de Fermat (1601-1665), afirma que esa ecuación no tiene solución natural.[8] Aunque abogado de profesión (y contemporáneo de Descartes), Fermat era el mejor matemático de su época. Dijo tener "una demostración verdaderamente maravillosa" de su afirmación, pero que no cabía en el estrecho margen del libro. Asómbrese usted: hasta hoy nadie ha sido capaz de reconstruir tal demostración ni, tampoco, de encontrar un solo contraejemplo a la afirmación de Fermat.

Es evidente que, una vez *dada* la cuádrupla de números (*w*, *x*, *y*, *z*.) es una simple cuestión de cálculo el decidir si se cumple o no la ecuación. Entonces podríamos imaginar un algoritmo para computadora que recorra todas las cuádruplas de números una tras otra y se pare cuando una satisfaga la ecuación. (Hemos visto que se pueden codificar en una sola cinta conjuntos finitos de números, de forma computable, esto es, simplemente como números individuales; de modo que podemos recorrer todas las cuádruplas siguiendo simplemente el orden natural de estos números individuales.) Si pudiéramos establecer que este algoritmo *no* se detiene nunca, entonces tendríamos una demostración de la tesis de Fermat.

De modo análogo, es posible referirnos en términos de una máquina de Turing que se detenga o no, a muchos otros problemas matemáticos no resueltos. Un ejemplo es la "conjetura de Goldbach", que afirma que cualquier número par mayor que 2 es la suma de dos números primos.[**]

Decidir si un número natural dado es primo o no es un proceso algorítmico, puesto que basta con comprobar su divisibilidad por números *menores* que él mismo, una cuestión que sólo requiere un cálculo *finito*. De la misma manera, podemos imaginar una máquina de Turing que recorra los números pares 6, 8, 10, 12, 14,... busque todas las formas diferentes de descomponerlos en pares de números nones

$$6 = 3 + 3, \quad 8 = 3 + 5, \quad 10 = 3 + 7 = 5 + 5, \quad 12 = 5 + 7,$$
$$14 = 3 + 11 = 7 + 7, \ldots$$

y verifique que *cada uno* de estos números pares se descomponga en *alguna* pareja de números primos. (Evidentemente no necesitamos comprobar parejas de sumandos pares, excepto 2 + 2, puesto que todos los primos, salvo el 2, son impares.) Nuestra máquina se detendrá sólo cuando llegue a un número par para el que *ninguno* de los pares de números en los que se descompone conste de dos primos. En tal caso tendríamos un contraejemplo a la conjetura de Goldbach, a saber, un número par (mayor que 2) que no es la suma de dos primos. Por lo tanto, si pudiéramos

[*] Recuérdese que por números naturales entendemos 0, 1,2, 3, 4, 5, 6,... La razón de escribir la ecuación en términos de x+1 y w+3, etc., en lugar de la forma más familiar $x^w + y^w = z^w$;x,y,z>0, w>2) del enunciado de Fermat, es que estamos suponiendo que x, w, etc., pueden ser cualquier número natural, comenzando por el cero.

[8] Para una discusión no técnica de las cuestiones relacionadas con este famoso enunciado, véase Devlin (1988),

[**] Recordemos que los números *primos* 2, 3, 5, 7, 11, 13, 17,... son aquellos números naturales que sólo son divisibles por sí mismos y por la unidad. Ni el O ni el 1 se consideran primos.

decidir si esta máquina de Turing se detendrá o no tendríamos una forma de decidir también la verdad o la falsedad de la conjetura de Goldbach.

Surge naturalmente una pregunta: ¿cómo decidir si una máquina de Turing particular (a la que se le introduce algún *input* específico) se detendrá o no en algún momento dado? Para muchas máquinas de Turing esto no sería difícil de responder; pero, en ocasiones, como hemos visto más arriba, la respuesta podría implicar la solución de algún otro problema matemático irresuelto. Por consiguiente, ¿existe algún procedimiento *algorítmico* que responda la cuestión que nos ocupa, el problema de la máquina que se detiene, de forma completamente automática? Turing demostró que no existe.

Su argumento es esencialmente el siguiente. Primero supongamos que, por el contrario, *sí* existe tal algoritmo.* En tal caso debe haber alguna máquina de Turing *H* que "decida" si la n-ésima máquina de Turing, al actuar sobre el número *m*, se detendrá o no. Digamos que su *output* es la cinta numerada O si no se para y 1 si lo hace:

$$H(n;m) = \begin{cases} 0 \text{ si } T_n(m) = \square \\ 1 \text{ si } T_n(m) \text{ se para} \end{cases}$$

Aquí podríamos hacer que la codificación del par (n,m) siguiera la misma regla que adoptamos para la máquina universal *U*. Sin embargo, esto tropezaría con el problema técnico de que para algunos números n(*v.g.* $n = 7$) T_n no está correctamente especificada y el marcador 11111 sería inadecuado para separar la *n* de la *m* en la cinta. Para evitar este problema supongamos que la *n* se codifica usando la notación binaria *expandida* en lugar de la simple notación binaria, y la m en forma binaria ordinaria, como antes. Entonces el marcador 110 será suficiente para separar la n de la m. El uso del punto y coma en $H(n;m)$, a diferencia de la *coma* en $U(n, m)$, indicará este cambio.

Imaginemos ahora una matriz infinita que enlista todos los *outputs* de todas las posibles máquinas de Turing que actúan sobre todos los diferentes *inputs* posibles. La n-ésima fila de la matriz muestra el *output* de la n-ésima máquina de Turing cuando *se* aplica a los diversos *inputs* 0, 1,2,3,4,...

* Este es un procedimiento matemático familiar —y potente— conocido como reducción al absurdo, en el que se supone que lo que se trata de demostrar es falso; si de ello se deriva una contradicción, queda establecido que el resultado es realmente *verdadero.*

m ,	0	1	2	3	4	5	6	7	8	...
n										
‰										
0	□	□	□	□	□	□	□	□	□	...
1	0	0	0	0	0	0	0	0	0	...
2	1	1	1	1	1	1	1	1	1	...
3	0	2	0	2	0	2	0	2	0	...
4	1	1	1	1	1	1	1	1	1	...
5	0	□	0	□	0	□	0	□	0	...
6	0	□	1	□	2	□	3	□	4	...
7	0	1	2	3	4	5	6	7	8	...
8	□	1	□	□	1	□	□	□	1	...
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	
1	2	3	5	7	11	13	17	19	23	...
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	

En esta tabla he hecho una pequeña trampa: no he enlistado las máquinas de Turing tal como están numeradas *realmente*. De haberlo hecho así hubiera dado lugar a una lista muy engorrosa puesto que todas las máquinas para las que n es menor que 11 no dan mas que □'s, y para n = 11 sólo dan 0's. Para hacer la lista más interesante desde el principio he supuesto que se hubiera conseguido elaborar una codificación mucho más eficaz. En realidad sólo he formado las entradas aleatoriamente, para dar una idea de cuál podría ser su apariencia general.

No estoy pensando realmente en *calcular* esta matriz mediante algún algoritmo. (De hecho no existe tal algoritmo, como veremos en un momento.) Simplemente *imaginemos* que la *verdadera* lista ha sido elaborada..., a lo mejor por Dios. Lo que nos causaría dificultades, si intentáramos calcular la matriz, sería la aparición de □'s pues no podríamos saber con certeza cuando colocar un □ en algún lugar ya que los cálculos continuarían indefinidamente.

Sin embargo, podríamos crear un procedimiento de cálculo para generar la tabla si se nos permitiera utilizar nuestra supuesta H, pues H nos diría dónde aparecen realmente los □'s. En lugar de esto, utilicemos H para eliminar todos los □ reemplazándolos con 0's. Esto se consigue precediendo la actuación de T_n sobre m por el cálculo $H(n;m)$; a continuación permitiremos que T_n actúe sobre m sólo si $H(n;m) = 1$ (esto es, sólo si el cálculo $T_n(m)$ da realmente una respuesta), y escribiendo simplemente 0 si $H(n;m) = 0$ (esto es, si $T_n(m)$ = □). Podemos escribir nuestro

nuevo procedimiento (es decir, el obtenido precediendo T_n (*m*) por la actuación de *H*(*n*;*m*)) en la forma

$$T_n(m) \times H(n; m).$$

(Aquí estoy usando un criterio común sobre el orden de las operaciones matemáticas; la que figura a la *derecha* debe ser la *primera* en realizarse. Nótese que, simbólicamente, tenemos □ x 0 = 0.) La tabla correspondiente ahora será

m ,	0	1	2	3	4	5	6	7	8	...
n										
%										
0	0	0	0	0	0	0	0	0	0	...
1	0	0	0	0	0	0	0	0	0	...
2	1	1	1	1	1	1	1	1	1	...
3	0	2	0	2	0	2	0	2	0	...
4	1	1	1	1	1	1	1	1	1	...
5	0	0	0	0	0	0	0	0	0	...
6	0	0	1	0	2	0	3	0	4	...
7	0	1	2	3	4	5	6	7	8	...
8	0	1	0	0	1	0	0	0	1	...
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	
197.	2	3	5	7	11	13	17	19	23	...
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	

Si suponemos que *H* existe, las filas de esta tabla constan de *secuencias computables*. (Entiendo por secuencia computable una sucesión infinita de números cuyos valores pueden ser generados por un algoritmo; es decir, que existe alguna máquina de Turing que, cuando se aplica a los números naturales *m* = 0, 1, 2, 3, 4, 5,..., uno por uno, da los términos de la secuencia.) Ahora tomemos nota de dos hechos acerca de esta tabla. En primer lugar, *toda* secuencia computable de números naturales debe aparecer en alguna parte (quizá muchas veces) entre sus filas. Esta propiedad era ya válida para la tabla original con sus □s. Simplemente hemos *añadido* algunas filas para reemplazar las máquinas de Turing "inútiles" (esto es, las que producen al menos un □). En segundo lugar, al hacer la suposición de que la máquina de Turing *H* existe realmente, la tabla ha sido *generada de manera computable* (esto es, generada mediante algún algoritmo definido), a saber, mediante el procedimiento $T_n(m) \times H(n; m)$. Es decir, existe alguna máquina de Turing Q que cuando actúa sobre el par de números (*n*, *m*) produce la entrada apropiada en la

tabla. Para esto, podemos codificar n y m en la cinta de Q de la misma forma que lo hacíamos para *H*, y tendremos

$$Q(n; m) = T_n(m) \times H(n; m).$$

Apliquemos ahora una variante de un ingenioso y potente artificio, el "corte diagonal" de Georg Cantor. (Veremos la versión original del corte diagonal de Cantor en el próximo capítulo.) Consideremos los elementos de la diagonal principal, marcados con guarismos en negritas:

0	0	0	0	0	0	0	0	0 ...
0	**0**	0	0	0	0	0	0	0...
1	1	**1**	1	1	1	1	1	1 ...
0	2	0	**2**	0	2	0	2	0 ...
1	1	1	1	**1**	1	1	1	1 ...
0	0	0	0	0	**0**	0	0	0...
0	0	1	0	2	0	**3**	0	4 ...
0	1	2	3	4	5	6	**7**	8 ...
0	1	0	0	1	0	0	0	**1 ...**
.	.		.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

Estos elementos forman cierta secuencia 0, 0, 1, 2, 1, 0, 3, 7, 1,.cada uno de cuyos términos sumamos 1:

1, 1, 2, 3, 2, 1, 4, 8 ,2....

Este es claramente un procedimiento computable y, dado que nuestra tabla estaba generada de manera computable, nos proporciona una nueva secuencia también computable, de hecho la secuencia $1 + Q(n;n)$, esto es,

$$1 + T_n(n) \times H(n;n)$$

(puesto que la diagonal se obtiene haciendo *m* igual a *n*). Ahora bien, nuestra tabla contiene *todas* las secuencias computables, de modo que nuestra nueva secuencia debe estar en alguna parte de la lista. Pero esto es imposible. En efecto, nuestra nueva secuencia difiere de la primera fila en el primer dígito, de la segunda fila en el segundo dígito, de la tercera fila en el tercer dígito, y así sucesivamente. Esto es una contradicción manifiesta. Es esta contradicción la que establece lo que hemos estado tratando de demostrar, a saber: que en realidad no existe la máquina de Turing *H*. *No existe un algoritmo universal para decidir si una máquina de Turing se detendrá o no.*

Otra manera de *parafrasear* este argumento consiste en señalar que, suponiendo que *H* existe, hay algún número de máquina de Turing, digamos *k*, para el algoritmo $1 + Q(n; n)$, de modo que tenemos

$$1 + T_n(n) \times H(n;n) = T_k(n)$$

Pero si sustituimos $n = k$ en esta relación obtenemos

$$1 + T_k(k) \times H(k;k) = T_k(k)$$

Y esto es una contradicción porque si $T_k(k)$ se detuviera obtendríamos la relación imposible

$$1+T_k(k) = T_k(k)'$$

(puesto que $H(k; k) = 1$), mientras que si $T_k(k)$ no se detiene (de modo que $H(k;k) = 0$) tendremos la relación igualmente inconsistente

$$1+0 = \square .$$

El si una máquina de Turing concreta se detiene o no, es un problema matemático perfectamente bien definido (y ya hemos visto que, recíprocamente, varias cuestiones matemáticas importantes pueden ser enunciadas en términos de parada de máquinas de Turing). De este modo, al demostrar que no existe ningún algoritmo para decidir la cuestión de la parada de las máquinas de Turing, Turing demostró (como lo hiciera Church utilizando una aproximación bastante diferente) que no puede haber un algoritmo general para decidir cuestiones matemáticas. Concluimos entonces que el *Entscheidungsproblem* de Hilbert no tiene solución Esto no quiere decir que en ningún caso particular seamos capaces de decidir la verdad o la falsedad de alguna cuestión matemática concreta o decidir si una máquina de Turing dada se detendrá o no. Poniendo en juego el ingenio, o incluso el simple sentido común, podremos tal vez decidir semejante cuestión en un caso dado. (Por ejemplo, si la lista de instrucciones de una máquina de Turing no contiene *ninguna* orden ALTO, o contiene *solamente* órdenes ALTO, entonces el sólo sentido común es suficiente para decirnos si se detendrá o no.) Pero no hay ningún algoritmo que funcione para *todas* las cuestiones matemáticas, ni para *todas* las máquinas de Turing con todos los números sobre los que podrían actuar.

Pudiera parecer que con esto hemos establecido que hay, al menos, *algunas* cuestiones matemáticas indecidibles. Nada de eso. *No* hemos demostrado que exista alguna tabla de máquina de Turing especialmente complicada para la que, en sentido absoluto, sea imposible decidir si la máquina se detendrá o no cuando se le introduzca algún número especialmente molesto; más bien, casi hemos demostrado lo contrario, como veremos en un momento. No hemos dicho nada sobre la insolubilidad de problemas *particulares*, sino sólo sobre la insolubilidad *algorítmica* de *familias* de problemas. En cualquier caso particular, la respuesta es o "sí" o "no", de modo que ciertamente *habrá* un algoritmo para decidir este caso concreto, a saber: el algoritmo que simplemente dice "sí" cuando se le plantea el problema, o el que dice "no", según sea el caso. La dificultad estriba, por supuesto, en que no podemos saber *cuál* de los dos algoritmos usar. Este es el problema de decidir sobre la verdad matemática de un enunciado particular, y no el de la decisión sistemática para una familia de enunciados. Es importante darse cuenta de que los algoritmos no deciden, por sí mismos, sobre la verdad matemática. La *validez* de un algoritmo debe establecerse siempre por medios externos.

CÓMO GANARLE A UN ALGORITMO

Volveremos más tarde a la cuestión de decidir sobre la verdad de enunciados matemáticos, cuando tratemos el teorema de Gödel (véase el capítulo IV). De momento quiero señalar que el argumento de Turing es mucho más constructivo y mucho menos negativo de lo que haya podido

Parecer hasta ahora. Ciertamente *no* hemos mostrado ninguna máquina de Turing específica para la que, sea indecidible si se detendrá o no en sentido absoluto. En realidad, si examinamos cuidadosamente el razonamiento, descubrimos que nuestro propio método implícitamente *nos da la respuesta* para las máquinas "especialmente complicadas", en apariencia, que construimos utilizando el procedimiento de Turing.

Veamos de dónde sale esto. Supongamos que tenemos un algoritmo que *a veces* es efectivo para decir cuándo no se detendrá una máquina de Turing. El procedimiento de Turing, como se esbozó más arriba, mostrará *explícitamente* el cómputo de una máquina de Turing para el que ese algoritmo particular no será capaz de decidir si el cómputo se detiene o no. Sin embargo, al hacerlo nos permitirá conocer la respuesta de ese caso particular. El cómputo de la máquina de Turing que mostramos antes *no* se detendrá.

Para ver cómo funciona esto en detalle supongamos que tenemos un algoritmo así, que es a veces efectivo. Como antes, denotemos este algoritmo (máquina de Turing) por H, pero ahora permitamos que el algoritmo pueda no estar en lo cierto al decirnos si una máquina de Turing se detendrá:

$$H(n;m) = \begin{cases} 0 \text{ o } \square & \text{si } T_n(m) = \square \\ 1 & \text{si } T_n(m) \text{ se para} \end{cases}$$

de modo que hay la posibilidad de que $H(n;m) = \square$ cuando $T_n(m) = \square$. Existen muchos de estos algoritmos $H(n;m)$. (Por ejemplo, el que simplemente produzca un 1 en cuanto $T_n(m)$ se detenga, aunque *este* algoritmo particular no sería de mucha utilidad que digamos).

Podemos completar el procedimiento de Turing igual que antes, excepto que en lugar de reemplazar *todos* los $\square$s por 0's dejamos ahora algunos $\square$s sin cambiar. Como antes, nuestro procedimiento diagonal nos dará

$$1 + T_n(n) \times H(n;n)$$

como el n-ésimo término de la diagonal. (Obtendremos un $\square$ cada vez que $H(m; n) = \square$. Nótese que $\square \times \square = \square$, $1 + \square = \square$.) Esta es una operación impecable, de la manera en que es llevada a cabo por alguna máquina de Turing, digamos la k-ésima, y así *tenemos*

Consideremos el k-ésimo término diagonal, esto es $n = k$, *y* obtendremos

$$1 + T_k(k) \times H(k;k) = T_k(k).$$

Si el cómputo $T_k(k)$ se detiene tendremos una contradicción (puesto que se supone que $H(k;k)$ es 1 cuando $T_k(k)$ se detiene, y la ecuación entonces es inconsistente: $1 + T_k(k) = T_k(k)$). Así que $T_k(k)$ no puede detenerse, es decir

$$T_k(k) = \square$$

Pero el algoritmo no puede "saber" esto, ya que si diera $H(k;k) = 0$ tendríamos de nuevo una contradicción (simbólicamente, tendríamos la relación no válida: $1 + 0 = \square$).

Por consiguiente, si podemos hallar k sabremos cómo construir nuestro cálculo específico para derrotar al algoritmo, ¡pero para el que *nosotros* sabemos la respuesta! ¿Cómo hallar k? Esta es

difícil tarea. Lo que tenemos que hacer es observar en detalle la construcción de $H(n; m)$ y de $T_n(m)$ y luego ver en detalle cómo actúa $1 + T_n(n) \times H(n;n)$ como máquina de Turing. Hallamos el número de esta máquina de Turing, que es k. Esto sería ciertamente complicado de llevar a cabo en detalle pero puede hacerse.* Debido a su complejidad no estaríamos en absoluto interesados en el cálculo $T_k(k)$, de no ser por el hecho de que lo hemos producido especialmente para derrotar al algoritmo H. Lo importante es que tenemos un procedimiento bien definido, cualquiera que sea el H que nos den, para hallar el correspondiente k para el que nosotros sabemos que $T_k(k)$ derrota a H, y para el que podemos superar al algoritmo. ¡Quizá nos consuele un poco el pensar que somos mejores que los simples algoritmos!

De hecho, el procedimiento está tan bien definido que podríamos hallar un *algoritmo* para generar k, dado H. Por lo tanto, antes de quedarnos demasiado satisfechos, tenemos que darnos cuenta de lo que *este* algoritmo puede mejorar[9] en H, ya que, en efecto, él "sabe" que $T_k(k) = \square$ ¿O no lo sabe? Nos ha sido de ayuda en la descripción anterior el utilizar el término antropomorfo "saber" con referencia a un algoritmo. Sin embargo, ¿no somos *nosotros* quienes estamos "sabiendo", mientras que el algoritmo tan sólo sigue las reglas que le hemos dicho que siga? ¿O estamos nosotros mismos simplemente siguiendo reglas para cuyo seguimiento estamos programados por la construcción de nuestros cerebros y de nuestro entorno? Este asunto no es simplemente una cuestión de algoritmos sino también una cuestión acerca de cómo juzgamos lo que es cierto y lo que es falso. Estos son temas capitales a los que tendré que volver más adelante. La cuestión de la verdad matemática (y su naturaleza no algorítmica) será estudiada en el capítulo IV. Cuando menos, ahora tendremos alguna noción sobre los *significados* de los términos "algoritmo" y "computabilidad", y una comprensión de algunos de los temas relacionados.

EL CÁLCULO LAMBDA DE CHURCH

El concepto de computabilidad es una idea matemática muy importante y bella. Es también notablemente reciente —si se tiene en cuenta su naturaleza fundamental para las matemáticas—, ya que fue desarrollado por primera vez en los años treinta. Es una idea que atraviesa *todas* las áreas de la matemática (aunque es bastante cierto que la mayoría de los matemáticos, por ahora, no se preocupa muy a menudo por cuestiones de computabilidad). El potencial de la idea reside en parte en el hecho de que algunas operaciones bien definidas en matemáticas *no* son computables (como el detenerse, o no, de una máquina de Turing; veremos otros ejemplos en el capítulo IV). En efecto, si no existieran estas cosas no computables el concepto de computabilidad no tendría mucho interés matemático. Los matemáticos, después de todo, aman los rompecabezas. Puede ser un rompecabezas intrigante para ellos decidir, en relación con alguna operación matemática, si es computable o no. Es especialmente intrigante debido a que la solución general a dicho rompecabezas es en sí misma no computable.

---

* De hecho, ya se ha conseguido lo más difícil mediante la construcción anterior de la máquina universal de Turing U, puesto que ello nos posibilita escribir $T_n(n)$ como una máquina de Turing actuando sobre n.

9 Por supuesto, también podríamos derrotar a este algoritmo mejorado aplicando simplemente el procedimiento precedente una vez más. Podemos entonces utilizar este nuevo conocimiento para mejorar aún más nuestro algoritmo; pero también podríamos derrotar a este otro, y así sucesivamente. El tipo de consideración a que nos lleva este procedimiento iterativo será discutido en relación con el teorema de Gödel, en el capítulo IV.

Una cosa debe quedar clara: el concepto de computabilidad es un concepto matemático verdaderamente "absoluto". Es una idea abstracta que queda más allá de cualquier realización concreta en términos de máquinas de Turing, tal como las he descrito. Como he resaltado antes, no necesitamos ligar ningún significado concreto a las "cintas" y "estados internos", etc., que caracterizan la ingeniosa aunque particular aproximación de Turing. Existen también otras formas de expresar la idea de computabilidad, de las que la primera históricamente fue el notable "cálculo lambda" del lógico estadounidense Alonzo Church con la ayuda de Stephen C. Kleene. El método de Church era bastante diferente e indudablemente más abstracto que el de Turing. En efecto, en la forma que Church estableció sus ideas, apenas existe una conexión obvia entre éstas y cualquier cosa que pudiéramos llamar "mecánica". La idea clave que subyace tras el método de Church es verdaderamente *abstracta* en su misma esencia: una operación matemática que Church, de hecho, llamaba "abstracción".

Creo que merece la pena dar una breve descripción del esquema de Church no sólo porque hace hincapié en que la computabilidad es una idea matemática, independiente de cualquier concepto particular sobre la máquina de computar, sino también porque ilustra la potencia de las ideas abstractas en matemáticas. El lector que no esté versado en ideas matemáticas ni intrigado por propio gusto por tales cosas, puede, en este punto, saltar al siguiente capítulo sin que haya una pérdida significativa en el curso del argumento. De todas formas, creo que tales lectores podrían obtener provecho acompañándome un poco más y siendo testigos de la mágica economía del esquema de Church (véase Church, 1941). En este esquema estamos interesados en un "universo" de objetos, denotados, por ejemplo por

$$a, b, c, d, \ldots, z, a', b', \ldots, z', a'', b'', \ldots, a''', \ldots. a'''', \ldots$$

cada uno de los cuales representa una operación matemática o *función.* (La *razón para* las letras con prima es simplemente permitir una cantidad ilimitada de símbolos para denotar tales funciones.) Los "argumentos" de estas funciones — es decir, los cosas sobre las que estas funciones actúan — son otras cosas del mismo tipo, esto es, también funciones. Además, el resultado (o "valor") de una función actuando sobre otra es de nuevo una función. (Hay ciertamente una maravillosa economía de conceptos en el sistema de Church) Así, cuando escribimos*:

$$a = bc$$

queremos decir que el resultado de la función b actuando sobre la función c es otra función a. No hay ninguna dificultad para expresar la idea de una función de dos o más variables en este esquema. Si queremos pensar en f como una función de dos variables p y q, por ejemplo, podemos escribir simplemente

$$(fp)q$$

(que es el resultado de la función fp aplicada a q). Para una función de tres variables consideramos

$$((fp)q)r,$$

* Una forma más familiar de notación hubiera consistido en escribir $a = b(c)$, por ejemplo, pero estos paréntesis no son realmente necesarios y es mejor que los omitamos. El incluirlos de forma consistente nos llevaría a fórmulas bastante farragosas, como $(f(p))(q)$ $y(f(p))(q))(r)$ en lugar de $(fp)q$ y $((fp)q)r$, respectivamente.

y así sucesivamente.

Ahora llega la poderosa operación de *abstracción*. Para ésta utilizamos la letra griega λ (lambda) seguida inmediatamente por una letra que representa una de las funciones de Church, digamos *x*, que consideramos como una "variable muda". Toda aparición de la variable *x* en la expresión entre corchetes que sigue inmediatamente es considerada simplemente como un "hueco" en el que se puede sustituir cualquier cosa que siga a la expresión completa. Así, si escribimos

$$\lambda x.[fx]$$

queremos expresar la función que cuando actúa sobre, pongamos por caso, *a* produce el resultado *fa*. Es decir

$$(x.[fx])a=fa$$

En otras palabras, $\lambda x.[fx]$ es simplemente la función *f*, es decir

$$\lambda x.[fx]=f$$

Esto requiere un poco de reflexión. Es una de esas sutilezas matemáticas que parece tan pedante y trivial a primera vista que uno tiende a pasar por alto el punto esencial. Consideremos un ejemplo tomado de las matemáticas familiares de la escuela. Sea la función *f* la operación matemática de tomar el seno de un ángulo, de modo que la función abstracta "sen" se define por

$$\lambda x.[sen x]=sen$$

(No se preocupe por cómo la "función" *x* puede ser considerada como un ángulo. Pronto veremos algo sobre cómo los números pueden ser tomados como funciones; y un ángulo es sólo un tipo de número.) Hasta aquí, esto es más bien trivial. Pero imaginemos que la notación "sen" no hubiera sido inventada, aunque conocemos el desarrollo en serie de potencias para sen *x*:

$$x-\frac{1}{6}x^3+\frac{1}{120}x^5-\ldots$$

Entonces podríamos definir

$$sen=\lambda x\left[x-\frac{1}{6}x^3+\frac{1}{120}x^5-\ldots\right]$$

Nótese que de forma aún mas simple, podríamos definir, pongamos por caso, la operación de "un sexto del cubo" para la que no hay notación "funcional" estándar:

$$Q=\lambda x.\left(\frac{1}{6}x^3\right)$$

y encontrar por ejemplo

$$Q(a+1)=1/6(a+1)^3=1/6ia^3+1/2a^2+1/2a+1/6$$

Más pertinentes para la discusión actual serían expresiones formadas simplemente a partir de las operaciones funcionales elementales de Church, tales como

$$\lambda f\left[f(fx)\right]$$

Esta es la función que, cuando actúa sobre otra función, digamos g, produce g iterada dos veces actuando sobre x, es decir

$$(\lambda f[f(fx)])g=g(gx)$$

También podríamos haber "abstraído fuera" primero la x para obtener

$$\lambda f.[\lambda x.[f(fx)]],$$

que podemos abreviar

$$\lambda fx[f(fx)]$$

Esta es la operación que, cuando actúa sobre g, produce la función "g iterada dos veces". De hecho ésta precisamente es la función que Church identifica con el número natural 2:

$$\mathbf{2} = \lambda fx.[f(fx)]$$

de modo que $(2g)y = g(gy)$. Análogamente define:

$$\mathbf{3} = \lambda fx.[f(f(fx))] \qquad \mathbf{4} = \lambda fx.[f(f(f(fx)))], \qquad \text{etc.,}$$

junto con

$$\mathbf{1} = \lambda fx.[fx], \qquad \mathbf{0} = \lambda fx.[x]$$

Realmente el "2" de Church se parece más a "doble" y su "3" a "triple", etc. Así, la acción de 3 sobre una función f, a saber $3f$, es la operación "Iterar f tres veces". La acción de $3f$ sobre y sería entonces $(\mathbf{3}f)y = f(f(f(y)))$

Veamos ahora cómo puede expresarse en el sistema de Church una operación aritmética muy sencilla, a saber, la operación de añadir 1 a un número. Definamos

$$\mathrm{S}=\lambda\, abc.[b((ab)c)].$$

Para ilustrar el hecho de que S añade simplemente 1 a un número descrito en la notación de Church, comprobémoslo en 3:

$$\mathrm{S}3 = \lambda\, abc.[b((ab)c)]3 = \lambda\, bc.[b((3b)c)] = \lambda\, bc.[b(b(b(bc)))] = 4,$$

puesto que $(3b)c = b(b(bc))$. Evidentemente esto también se aplica a cualquier otro número natural. (De hecho $\lambda\, abc.[(ab)(bc)]$ hubiera hecho lo mismo que S.)

¿Cómo sería la multiplicación de un número por dos? Esta duplicación se lograría mediante

$$\mathrm{D} = \lambda\, abc.[(ab)((ab)c)],$$

que de nuevo se ilustra por su actuación sobre 3:

$$\mathrm{D} = \lambda\, abc.[(ab)((ab)c)]3 = \lambda\, bc\, .[(3b)(3b)c)] = \lambda\, bc.[(3b)(b(b(bc)))]$$
$$= \lambda\, bc.[b(b(b(b(b(bc)))))] = 6.$$

De hecho, las operaciones aritméticas básicas de adición, multiplicación y elevación a una potencia pueden definirse, respectivamente, por:

$$\mathrm{A}=\lambda\, fgxy[((fx)(gx))y],$$

$$M=\lambda fgx[f(gx)],$$

$$P=\lambda fg[fg]$$

La lectora o el lector pueden ocuparse en convencerse por sí mismos — o por cualquier otro en quien tenga plena confianza— de que efectivamente

$$(A\mathbf{m})\mathbf{n} = \mathbf{m} + \mathbf{n}, \qquad (M\mathbf{m})\mathbf{n} = \mathbf{m} \times \mathbf{n}, \qquad (P\mathbf{m})\mathbf{n} = \mathbf{n}^{\mathbf{m}},$$

donde **m** y **n** son funciones de Church para dos números naturales, **m + n** es la función de Church para su suma, y así para las demás. La última de éstas es la más sorprendente. Comprobémosla para el caso **m** = 2, **n** = 3:

$$(P2)3=((\lambda fg[fg])2)3 = (\lambda g.[2g])3$$

$$=(\lambda g.[\lambda fx.[f(fx)]g])3= \lambda gx.[g(gx)]3$$

$$=\lambda x.[3(3x)]= \lambda x[\lambda fy[f(f(fy))](3x)]$$

$$=\lambda xy.[(3x)((3x)((3x)y))]$$

$$=\lambda xy.[(3x)((3x)(x(x(xy))))]$$

$$=\lambda xy.[(3x)(x(x(x(x(x(xy))))))]$$

$$=\lambda xy.[x(x(x(x(x(x(x(x(xy)))))))))] = \mathbf{9} = \mathbf{3^2}.$$

Las operaciones de substracción y división no se definen tan fácilmente (y necesitamos efectivamente algún convenio sobre lo que hacer con "m - n" cuando m es menor que n y con "m + n" cuando m no es divisible por n). De hecho, un hito importante en la materia sucedió a comienzos de los años treinta cuando Kleene descubrió la forma de expresar la operación de substracción dentro del esquema de Church. A continuación siguieron otras operaciones. Finalmente, en 1937, Church y Turing, independientemente, demostraron que cualquier operación computable (o algorítmica) — ahora en el sentido de las máquinas de Turing — puede lograrse en términos de una de las expresiones de Church (y viceversa).

Éste es un hecho verdaderamente notable, y sirve para subrayar el carácter fundamentalmente objetivo y matemático de la noción de computabilidad. La noción de computabilidad de Church tiene, a primera vista, muy poco que ver con las máquinas computadoras. Pese a todo tiene, en cualquier caso, ciertas relaciones fundamentales con la computación práctica. En particular, el potente y flexible lenguaje LISP incorpora, de un modo esencial, la estructura básica del cálculo de Church.

Como indiqué antes, hay también otras maneras de definir la noción de computabilidad. El concepto de Post de máquina computadora estaba muy próximo al de Turing y fue presentado independientemente y casi al mismo tiempo. Existía también una definición bastante más manejable de computabilidad (recursividad) debida a J. Herbrand y Gödel. H. B. Curry en 1929, y también M. Schónfinkel en 1924, tuvieron un poco antes una aproximación diferente, a partir de la que se desarrolló en parte el cálculo de Church (véase Gandy, 1988). Las aproximaciones modernas a la computabilidad (tales como la de una *máquina de registro ilimitado*, descrita en Cutland, 1980) difieren considerablemente en los detalles respecto a la original de Turing, y son bastante más prácticas. No obstante, el *concepto* de computabilidad sigue siendo el mismo, cualquiera que sea la aproximación que se adopte.

Al igual que sucede con muchas otras ideas matemáticas, en especial las más profundamente bellas y fundamentales, la idea de computabilidad parece tener una especie de *realidad platónica* autónoma. En los dos próximos capítulos tendremos que volver a esta misteriosa cuestión de la realidad platónica de los conceptos matemáticos en general.

III. MATEMÁTICA Y REALIDAD

LA TIERRA DE TOR'BLED-NAM

IMAGINEMOS QUE HEMOS REALIZADO UN LARGO VIAJE a algún mundo lejano. Llamaremos a este mundo Tor'Bled-Nam. Nuestros sensores de detección remota han captado una señal que se muestra ahora en una pantalla frente a nosotros. La imagen se focaliza y vemos (fig. III.1):

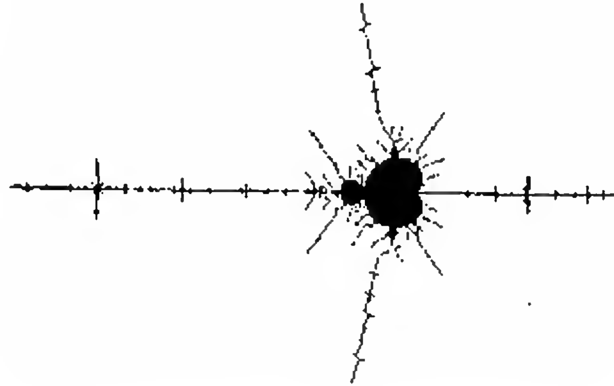


FIGURA III. 1. *Una primera impresión de un mundo extraño.*

¿Qué puede ser? ¿Es algún insecto de extraña apariencia? Tal vez, en lugar de ello, sea un lago de color oscuro con muchas cadenas montañosas que se adentran en él. ¿O podría ser una inmensa ciudad extraterrestre con una forma extraña y con carreteras que salen en varías direcciones hacia las pequeñas ciudades y pueblos vecinos? Quizá sea una isla; tratemos entonces de descubrir si existe algún continente próximo al que está asociada. Podemos hacer esto "retrocediendo", reduciendo la amplificación de nuestro dispositivo sensor en un factor lineal de aproximadamente quince. Dicho y hecho, el mundo completo surge a la vista (fig. III.2):

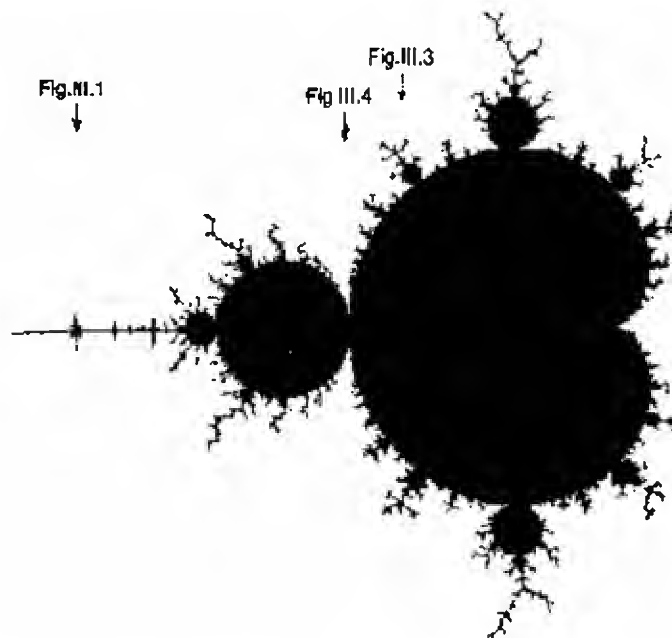


FIGURA III.2. *"Tor'Bled-Nam" en su totalidad. Las localizaciones de las ampliaciones que se muestran en las figs. III.1, III.3 y III.4 son las indicadas bajo las flechas.*

Nuestra "isla" se ve como un punto pequeño indicado bajo "fig. III.1" en la fig. III.2. Todos los filamentos (¿ríos, carreteras, puentes?) que parten de la isla original tienen un extremo final, con la excepción del que sale del interior de la grieta de su lado derecho, que acaba por unirse a un objeto mucho mayor que vemos representado en la fig. III.2. Este objeto mayor es claramente similar a la isla que vimos primero, aunque no es exactamente igual. Si examinamos con mayor atención lo que parece ser la línea costera de este objeto vemos innumerables protuberancias — redondas, si bien ellas mismas poseen sus propias protuberancias similares. Cada pequeña protuberancia parece estar unida a otra mayor en alguna región minúscula, dando lugar a una serie de verrugas sobre otras verrugas. A medida que la imagen se hace más nítida, vemos miríadas de pequeñísimos filamentos que emanan de la estructura. Los propios filamentos se ramifican en varios lugares y a menudo serpentean de forma imprevisible. En ciertos puntos de los filamentos parecen apreciarse pequeños nudos de más complejidad que nuestro dispositivo sensor, con su amplificación actual, no puede resolver. Evidentemente el objeto no es una isla o continente real, ni ningún tipo de paisaje. Tal vez, después de todo, estamos viendo algún monstruoso escarabajo, y lo primero que vimos era parte de su prole, unida todavía a él por algún tipo de cordón umbilical filamentoso.

Tratemos de examinar la naturaleza de una verruga de nuestra criatura, aumentando la amplificación de nuestro dispositivo sensor en un factor lineal de aproximadamente diez (fig. III.3 —cuya localización está indicada bajo "fig. III.3" en la fig. III.2). La propia verruga tiene un fuerte parecido con la criatura global —excepto sólo en el punto de unión—. Nótese que hay varios lugares en la fig. III.3 en donde se juntan cinco filamentos. Hay quizá una cierta "quinariedad" en esta verruga particular (como habría una "ternariedad" en la verruga superior). De hecho, si examináramos la siguiente verruga de tamaño apreciable, hacia la parte inferior izquierda en la fig. III.2, descubriríamos una "septenariedad" en ella; y una "nonariedad" en la siguiente, y así sucesivamente. A medida que nos adentramos en la grieta entre las dos regiones mayores de la fig. III.2, encontramos verrugas a la derecha caracterizadas por números impares que incrementan de dos en dos. Examinemos más profundamente esta grieta, aumentando la ampliación de la fig. III.2 en un factor de unos diez (fig. III.4).



FIGURA III.3. *Una verruga con una "quinariedad" en sus filamentos.*

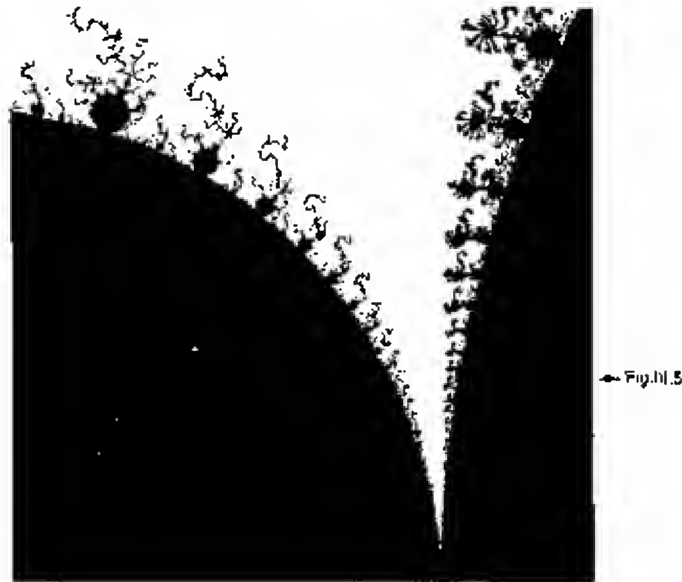


FIGURA III.4. *La grieta principal. El "Valle de los caballos de mar" es apenas perceptible en la parte inferior Derecha.*

Vemos otras numerosas verrugas minúsculas y también una agitada actividad. A la derecha, podemos distinguir algunas pequeñas "colas de caballo de mar" espirales —en un área que llamaremos "Valle de los caballos de mar"—. Aquí encontraremos, si la amplificación se aumenta lo suficiente, varias "anémonas de mar" o regiones con una apariencia floral diferente. Tal vez, después de todo, se trate realmente de alguna exótica línea costera —quizá algún arrecife coralino, con vida de todo tipo—. Lo que pudiera haber parecido una flor revelará, en una posterior ampliación, estar compuesto de miríadas de pequeñísimas aunque increíblemente complicadas estructuras, cada una con numerosos filamentos y colas espirales retorcidas. Examinemos con algún detalle una de las mayores colas de caballo de mar, a saber, la que es discernible donde se indica como "fig. III.5" en la fig. III.4 (que está unida a una verruga con una ¡"29-riedad"!). Con una posterior amplificación de 250 aumentos se nos presenta la espiral mostrada en la fig. III.5.

Descubrimos que ésta no es una cola ordinaria sino que ella misma está formada por más complicados remolinos hacia uno u otro lado, con innumerables espirales minúsculas y regiones que semejan pulpos y caballos de mar.

En muchos lugares la estructura está entrelazada precisamente donde se juntan dos espirales. Examinemos uno de estos lugares (indicado bajo "fig. III.6" en la fig. III.5) incrementando nuestra amplificación en un factor de aproximadamente treinta. ¡Ya está!; ¿distinguimos un objeto extraño, aunque ahora familiar, en el centro? Una amplificación posterior en un factor de seis (fig. III.7) revela una minúscula criatura bebé, ¡casi idéntica a la estructura global que hemos estado examinando! Si miramos más de cerca vemos que los filamentos que emanan de ella difieren un poco de los de la estructura principal, y se retuercen y extienden hasta distancias relativamente mucho mayores. Pero la propia criatura minúscula apenas parece diferir de su progenitor, hasta el grado de poseer incluso su propia prole en posiciones estrechamente análogas. Podríamos examinarlas una vez más si quisiésemos aumentando la amplificación. Los

nietos se parecerán también a su antepasado común, y ya nos sentimos dispuestos a creer que esto continúa indefinidamente.



FIGURA III.5. *Un primer plano de una cola de caballo de mar.*



FIGURA III. 6. *Una ampliación adicional de un punto de unión en el que se juntan dos espirales. Un pequeño bebé es apenas visible en el punto central.*

Podemos explorar este extraordinario mundo de Tor'Bled-Nam hasta donde deseemos, ajustando nuestro dispositivo sensor a mayores y mayores grados de amplificación. Encontramos una variedad sin fin: no hay dos regiones que sean exactamente iguales, pero hay un aire general al que pronto nos acostumbramos. La ahora familiar criatura escarabajoide emerge en escalas más y más pequeñas. En todo momento las estructuras filamentosas vecinas difieren de las que habíamos visto antes, y se nos presentan fantásticas escenas nuevas de increíble complejidad.

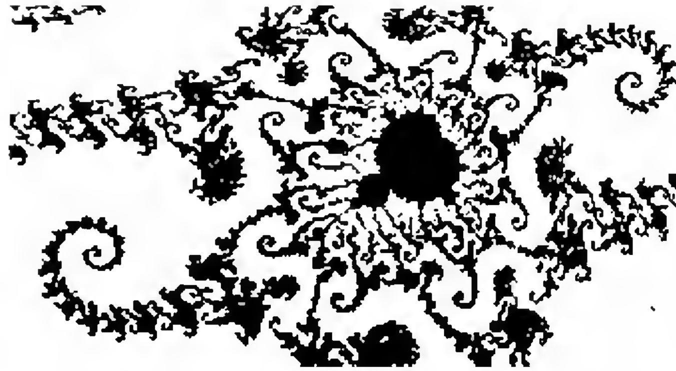


FIGURA III.7. *En una nueva ampliación se aprecia que el bebé se asemeja, mucho al mundo global.*

¿Qué es esta extraña, variada y maravillosamente intrincada tierra con la que hemos topado? Sin duda muchos lectores ya lo sabrán, pero otros no. Este mundo no es más que un objeto de la matemática abstracta: el conjunto conocido como conjunto de Mandelbrot.[1] Es complicado, sin duda; pero está generado por una regla de notable simplicidad. Para explicar la regla adecuadamente tendré que explicar antes lo que es un *número complejo*. Eso es lo que voy a hacer ahora. Necesitaremos los números complejos más adelante. Son absolutamente fundamentales para la estructura de la mecánica cuántica y son, por consiguiente, básicos para el funcionamiento del propio mundo en que vivimos. También constituyen uno de los grandes milagros de la matemática. Para explicar lo que es un número complejo necesitaré antes recordar al lector lo que significa el término "número real". Será útil, también, señalar la relación entre dicho concepto y la propia realidad del "mundo real".

NÚMEROS REALES

Recordemos que los números *naturales* son las cantidades enteras:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,...

Son los más elementales y básicos entre los diferentes tipos de números. Cualquier entidad discreta puede cuantificarse mediante el uso de números naturales: podemos hablar de veintisiete ovejas en un prado, de dos relámpagos, doce noches, mil palabras, cuatro conversaciones, cero ideas nuevas, un error, seis ausentes, dos cambios de dirección, etc. Los números naturales pueden sumarse o multiplicarse para dar nuevos números naturales. Son los objetos de nuestra discusión general de los algoritmos, como la que se dio en el capítulo precedente. Sin embargo, algunas operaciones importantes pueden llevarnos fuera del dominio de los números naturales; la resta es la más sencilla de éstas. Para definir la resta de una forma sistemática necesitamos los números *negativos*; podemos establecer, para este propósito, el sistema global de los *enteros*

..., -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7,...

[1] Véase Mandelbrot (1986). La particular secuencia de ampliaciones que he escogido está adaptada de las de Peitgen y Richter (1986), en donde se encontrarán muchas imágenes notablemente coloreadas del conjunto de Mandelbrot. Para más ilustraciones sorprendentes, véase Peitgen y Saupe (1988).

Ciertas cosas, como la carga eléctrica, los balances bancarios o las fechas* se cuantifican mediante números de este tipo. Pese a todo, estos números son aún de alcance demasiado limitado puesto que nos quedaríamos bloqueados cuando tratáramos de *dividir* un número por otro. En consecuencia, necesitaremos las *fracciones* o *números racionales* como son llamados

0, l,-l, 1/2, -1/2, 2,-2, 3/2, -3/2, 1/3,..

Esto es suficiente para las operaciones de la aritmética finita, pero para muchos otros propósitos necesitamos ir más allá e incluir operaciones infinitas o de paso al límite. Por ejemplo, la familiar — y de gran importancia en matemáticas — cantidad π aparece en muchas de estas expresiones infinitas. En particular, tenemos

$$\pi = 2[\left(\tfrac{2}{1}\right)\left(\tfrac{2}{3}\right)\left(\tfrac{4}{3}\right)\left(\tfrac{4}{5}\right)\left(\tfrac{6}{5}\right)\left(\tfrac{6}{7}\right)\left(\tfrac{8}{7}\right)\left(\tfrac{8}{9}\right)....]$$

$$\pi = 4(1-\tfrac{1}{3}+\tfrac{1}{5}-\tfrac{1}{7}+\tfrac{1}{9}-\tfrac{1}{11}....).$$

Estas son expresiones famosas, habiendo sido encontrada la primera por el matemático, gramático y experto criptógrafo inglés John Wallis en 1665; y la segunda por el matemático y astrónomo escocés (e inventor del primer telescopio reflector) James Gregory en 1671. Como sucede con π , los números definidos de esta forma *no* deben ser racionales (esto es, no de la forma *n*/*m*, en donde *n y m* son enteros con *m* distinto de cero). El sistema de números necesita ser *ampliado para* poder incluir tales cantidades.

Este sistema ampliado de números se denomina sistema de los números "reales", aquellos números familiares que pueden representarse como *expansiones decimales* infinitas, tales como:

-583.70264439121009538...

En términos de una representación semejante tenemos la bien conocida expresión para *n*:

$$\pi = 3.14159265358979323846...$$

Entre los tipos de números que también pueden representarse de este modo están las raíces cuadradas (o las raíces cúbicas o las raíces cuartas. etc.) de números racionales positivos, tales como:

$$\sqrt{2} = 1.414221356237309504...$$

o, de hecho, la raíz cuadrada (o raíz cúbica, etc.) de cualquier número real positivo, como sucede con la expresión para π encontrada por el gran matemático suizo Leonhard Euler:

$$\pi = \sqrt{[6(1+1/4+1/9+1/25+1/36+...)]}$$

Los números reales son, en efecto, el tipo familiar de números con los que tenemos que trabajar en la vida de cada día, aunque normalmente estamos interesados sólo en aproximaciones a tales números y preferimos trabajar con expansiones que incluyen sólo algunas cifras decimales. Sin embargo, en las proposiciones matemáticas los números reales tienen que ser especificados *exactamente*, y necesitamos algún tipo de descripción infinita como pueda ser una completa expansión decimal infinita, o quizá alguna otra expresión matemática infinita como las fórmulas anteriores para *n* dadas por Wallis, Gregory o Euler. (Normalmente utilizaré expansiones decimales en mis descripciones, pero sólo porque resultan más familiares. Para un matemático

* En realidad, las convenciones usuales sobre fechas no se ajustan totalmente a esto, ya que se omite el año cero.

hay varias maneras bastante más satisfactorias de presentar los números reales pero no necesitamos preocuparnos por eso ahora.)

Podría dar la impresión de que es imposible considerar una expansión infinita *completa* pero no es así. Un ejemplo sencillo en el que podemos considerar claramente la secuencia completa es

$$\tfrac{1}{3} = 0.3333333333333333...,$$

donde los puntos nos indican que la serie de los 3s continúa indefinidamente. Para considerar esta expansión, todo lo que necesitamos saber es que la expansión continúa de la misma forma indefinidamente con 3s. Todo número racional tiene una expansión decimal repetida (o finita), tal como

$$\tfrac{93}{74} = 1.2567567567567567567...,$$

donde la secuencia 567 se repite indefinidamente, y ésta también puede ser considerada en su totalidad. También, la expresión

0.220002222000002222220000000222222220..., v

que define un número *irracional*, puede ser considerada en su totalidad (las cadenas de 0s y 2s incrementan su longitud en uno cada vez) y pueden darse muchos ejemplos similares. En cada caso nos damos por satisfechos cuando conocemos una regla con arreglo a la que se construye la expansión. Si hay algún algoritmo que genera los sucesivos dígitos, el conocimiento de dicho algoritmo nos proporciona una forma de considerar toda la expansión decimal infinita. Los números reales cuya expansión puede ser generada mediante algoritmos se llaman números computables . (El uso de una expansión decimal en lugar de, pongamos por caso, una binaria no tiene importancia aquí. Los números que son "computables" en este sentido son los mismos cualquiera que sea la base utilizada para la expansión.) Los números reales π *y* $\sqrt{2}$ que hemos estado considerando son ejemplos de números computables. En ambos casos sería un poco complicado establecer la regla en detalle, pero no hay dificultad de principio.

Sin embargo, existen también muchos números reales que *no* son computables en este sentido. Hemos visto en el último capítulo que existen secuencias no computables que están, a pesar de todo, perfectamente bien definidas. Por ejemplo, podríamos tomar la expansión decimal cuyo *n*-ésimo dígito es 1 o 0 según se detenga o no la *n*-ésima máquina de Turing actuando sobre el número *n*. En general sólo pedimos que para un número real haya *cierta* expansión decimal infinita. No pedimos que exista un algoritmo para generar el *n*-ésimo dígito, ni siquiera que conozcamos algún tipo de regla que en principio defina cuál es realmente el *n* -ésimo dígito.[2] El trabajo con los números computables es cosa muy difícil. No se puede hacer que todas las operaciones sean computables aun cuando se trabaje sólo con números computables. Por ejemplo, ni siquiera es un asunto computable el decidir, en general, si dos números computables son iguales o no. Por razones de este tipo, es preferible trabajar, en su lugar, con *todos* los números reales, para los que las expansiones decimales pueden ser cualesquiera y no necesitan ser, pongamos por caso, una secuencia computable.

---

[2] Por lo que yo conozco, es un punto de vista consistente, aunque poco convencional, el exigir que hubiera siempre algún tipo de regla que determine cuál es realmente el n-ésimo dígito para un número real arbitrario, aunque esa regla pueda no ser efectiva ni siquiera definible en un sistema formal preasignado (véase el capítulo IV). Yo espero que sea consistente, puesto que es el punto de vista al que más me gustaría adherirme.

Señalaré finalmente que existe una identificación entre un número real cuya expansión decimal termina con una sucesión infinita de 9s y otro cuya expansión termina con una sucesión infinita de 0s; por ejemplo

$$-27.1860999999... = -27.1861000000...$$

¿CUÁNTOS NÚMEROS REALES HAY?

Hagamos una pequeña pausa para apreciar el alcance de la generalización que se ha conseguido al pasar de los números racionales a los números reales.

A primera vista se podría pensar que el número de enteros es mayor que el número de números naturales: todo número natural es un entero mientras que algunos enteros (a saber, los negativos) no son números naturales; y, análogamente, se podría pensar que el número de fracciones es mayor que el de enteros. Sin embargo, esto no es así. Según la potente y bella teoría de los números infinitos o transfinitos propuesta a finales del siglo XIX por el muy original matemático ruso-germano Georg Cantor, el número total de fracciones, el número total de enteros y el número total de números naturales son todos ellos el *mismo* número transfinito, denotado por $\aleph_0$ ("aleph subcero"). (Curiosamente una idea de este tipo había sido parcialmente anticipada unos 250 años antes, a comienzos del siglo XVII, por el gran físico y astrónomo italiano Galileo Galilei. Recordaremos algunos otros logros de Galileo en el capítulo V.) Se puede ver que el número de enteros es el mismo que el de naturales si establecemos una "correspondencia uno-a-uno" de la siguiente forma:

Enteros		**Numeros naturales**
0	↔	0
-1	↔	1
1	↔	2
-2	↔	3
2	↔	4
-3	↔	5
3	↔	6
-4	↔	7
.		.
.		.
.		.
$-n$	↔	$2n$-1
n	↔	$2n$
.		.
.		.
.		.

Nótese que cada entero (en la columna izquierda) y cada número natural (en la columna derecha) aparecen una y sólo una vez en la lista. La existencia de una correspondencia uno-a-uno como ésta es lo que establece, en la teoría de Cantor, que el número de objetos en la columna izquierda es el mismo que el número de objetos en la columna derecha. Por consiguiente, el número de enteros es el mismo que el de números naturales. En este caso el número es infinito, pero no

importa. (La única peculiaridad que ocurre con los números infinitos es que podemos olvidar algunos de los miembros de una lista ¡y *todavía* encontrar una correspondencia uno-a-uno entre las dos listas!) De manera análoga, aunque algo más complicada, podemos establecer una correspondencia uno-a-uno entre las fracciones y los enteros. (Para ello podemos adoptar una de las formas de representar pares de números naturales, numerador y denominador, como números naturales simples; véase capítulo II. Los conjuntos que pueden ser puestos en correspondencia uno-a-uno con los números naturales se llaman *numerables*, de modo que los conjuntos infinitos numerables tienen $\aleph_0$ elementos. Hemos visto así que los enteros son numerables y también lo son todas las fracciones.

¿Existen conjuntos que sean *no* numerables? Aunque hemos ampliado el sistema, pasando primero de los números naturales a los enteros, y luego a los números racionales, no hemos incrementado realmente el número de objetos con los que trabajamos. Hemos visto que el número de objetos es numerable en cada caso. Quizá el lector sacó de esto la impresión de que *todos* los conjuntos infinitos son numerables. No es así; la situación es muy diferente al pasar a los números reales. Uno de los logros notables de Cantor fue demostrar que realmente hay *más* números reales que racionales. El argumento que utilizó Cantor fue el del "corte diagonal" a que nos referimos en el capítulo II y que Turing adaptó en su argumento para demostrar que el problema de la detención de las máquinas de Turing es insoluble. El argumento de Cantor, como el posterior de Turing, procede por *reductio ad absurdum*. Supongamos que el resultado que estamos tratando de establecer es falso, es decir, que el conjunto de todos los números reales es numerable. Entonces, los números reales comprendidos entre 0 y 1 son ciertamente numerables, y tendremos *alguna* lista que proporciona una correspondencia uno-a-uno que empareja todos estos números con los números naturales, tal como:

Números Naturales		**Numeros Reales**
0	↔	0.**1**0357627183...
1	↔	0.1**4**329806115...
2	↔	0.02**1**66095213...
3	↔	0.430**0**5357779...
4	↔	0.9255**0**489101...
5	↔	0.59210**3**43297...
6	↔	0.636679**1**0457...
7	↔	0.8705007**4**193...
8	↔	0.04311737**8**04...
9	↔	0.786350811**5**0...
10	↔	0.4091673889**1**...

He marcado en negrita los dígitos de la diagonal. Para este listado particular, estos dígitos son

1,4,1,0,0,3,1,4,8,5,1,...

y el procedimiento del corte diagonal consiste en construir un número real (entre 0 y 1) cuya expansión decimal (tras el punto decimal) difiere de estos dígitos en cada uno de los lugares

correspondientes. Para poner un ejemplo concreto, digamos que el dígito será 1 donde quiera que el dígito de la diagonal es diferente de 1, y será 2 donde quiera que el dígito de la diagonal es 1. Por lo tanto, en este caso tenemos el número real

0.21211121112...

Este número real no puede figurar en nuestro listado puesto que difiere del primer número en la primera cifra decimal, del segundo número en la segunda cifra, del tercer número en la tercera cifra, etc. Existe una contradicción, ya que se suponía que nuestra lista contiene *todos* los números reales entre 0 y 1. Esta contradicción establece lo que estamos tratando de probar, a saber: que *no* existe correspondencia uno-a-uno entre los números reales y los números naturales y, en consecuencia, que el número de números reales es realmente *mayor* que el número de números racionales y *no* es numerable.

El número de números reales es el número transfinito llamado *C*. (*C* significa en este caso el *continuo*, otro nombre para el sistema de los números reales.) Podríamos preguntar por qué este número no es llamado $\aleph_1$, por ejemplo. En realidad, el símbolo $\aleph_1$ representa el siguiente número transfinito mayor que $\aleph_0$, y el decidir si efectivamente $C = \aleph_1$ constituye un famoso problema no resuelto, la llamada *hipótesis* del *continuo*.

Puede señalarse que, por el contrario, los números *computables* son numerables. Para numerarlos basta con hacer una lista, en orden numérico, de las máquinas de Turing que generan números reales (es decir, que producen los sucesivos dígitos de los números reales). Nos gustaría tachar de la lista cualquier máquina de Turing que genere un número real que ya haya aparecido antes en la lista. Puesto que las máquinas de

Turing son numerables debe darse ciertamente el caso de que los números reales computables son numerables. ¿Por qué no podemos utilizar el corte diagonal en esta lista y producir un nuevo número computable que *no* esté en la lista? La respuesta está en el hecho de que, en general, no podemos decidir computablemente si una máquina de Turing estará o no realmente en la lista. Eso supondría, en efecto, que fuéramos capaces de resolver el problema de la detención. Algunas máquinas de Turing pueden empezar a producir los dígitos de un número real, y luego quedarse en suspenso y no producir ningún otro dígito (debido a que "no se paran"). No hay medio computable de decidir qué máquinas de Turing se bloquearan de esta forma. Éste es básicamente el problema de la detención. Por lo tanto, aunque nuestro procedimiento del corte diagonal producirá algún número real, ese número no será un número computable. En realidad, podríamos haber utilizado este argumento para *demostrar* la existencia de números no computables. El argumento de Turing para demostrar la existencia de clases de problemas que no pueden resolverse algorítmicamente, como fue expuesto en el último capítulo, sigue precisamente esta línea de razonamiento. Veremos después otras aplicaciones del corte diagonal.

"REALIDAD" DE LOS NÚMEROS REALES

Dejando aparte la cuestión de la computabilidad, los números reales se llaman "reales" debido a que parecen proporcionar las magnitudes necesarias para la medida de distancias, ángulos, tiempo, energía, temperatura u otras numerosas cantidades geométricas y físicas. Sin embargo, la relación entre los números "reales" abstractamente definidos y las cantidades físicas no es tan nítida como uno pudiera imaginar. Los números reales se refieren a una *idealización matemática*

más que a cualquier cantidad física real objetiva. El sistema de los números reales tiene, por ejemplo, la propiedad de que entre dos cualesquiera de ellos, por muy próximos que estén, hay un tercero. No es en absoluto evidente que las distancias o los tiempos físicos posean esa propiedad. Si continuamos dividiendo la distancia física entre dos puntos alcanzaremos finalmente escalas tan pequeñas que el mismo concepto de distancia, en sentido ordinario, deja de tener significado. Se espera que éste sea el caso en la escala de la "gravitación cuántica" de una 10^{20}-ava parte del tamaño* de una partícula subatómica. Pero para reflejar los números reales tendríamos que ir a escalas infinitamente más pequeñas que esta: 10^{200} avo, 10^{2000}avo, o $10^{10^{200}}$ avo del tamaño de una partícula, por ejemplo. No es en absoluto evidente que tales escalas absurdamente ínfimas tengan ningún significado físico. Un comentario similar sería también válido para los correspondientemente minúsculos intervalos de tiempo.

El sistema de los números reales se escoge en física por su utilidad, simplicidad y elegancia *matemáticas*, junto con el hecho de que concuerda, en un rango muy amplio, con los conceptos físicos de distancia y tiempo. No se ha escogido porque sepamos que está de acuerdo con estos conceptos físicos en *todos* los rangos. Se podría esperar que no exista tal acuerdo a escalas muy pequeñas de distancia o tiempo. Es práctica común utilizar reglas para la medida de distancias simples, pero esas mismas reglas tendrán una naturaleza granular cuando descendamos a la escala de sus propios átomos. Esto, en sí mismo, no nos impide seguir utilizando los números reales de una forma aproximada, pero se necesita una sofisticación mucho mayor para la medida de distancias aún más pequeñas. Deberíamos tener al menos alguna sospecha de que pudiera haber en último término alguna dificultad de tipo fundamental para distancias en la escala más ínfima. Resulta que la naturaleza se muestra muy amable con nosotros y parece que los mismos números reales que nos hemos acostumbrado a utilizar para la descripción de las cosas a una escala cotidiana o mayor conservan su utilidad a escalas mucho más pequeñas que los átomos —con certeza por debajo de una centésima del diámetro "clásico" de una partícula subatómica, por ejemplo un electrón o un protón— y aparentemente por debajo de la "escala de la gravitación cuántica", ¡veinte órdenes de magnitud más pequeña que el tamaño de dicha partícula! Esta es una extraordinaria extrapolación a partir de la experiencia. El concepto familiar de distancia como número real parece ser válido también para el cuásar más remoto y aún más allá, lo que equivale a un rango global de al menos 10^{42}, y quizá 10^{60} o más. De hecho, esta adecuación del sistema de los números reales no se cuestiona normalmente. ¿Por qué se confía tanto en estos números para la exacta descripción de la física, cuando nuestra experiencia inicial de la importancia de tales números se reduce a un rango relativamente limitado? Esta confianza —quizá inmerecida— debe descansar (aunque este hecho no se reconoce a menudo) en la elegancia lógica, consistencia y potencia matemática del sistema de los números reales junto con una creencia en la profunda armonía matemática de la naturaleza.

NÚMEROS COMPLEJOS

El sistema de los números reales no tiene el monopolio de la potencia y elegancia matemáticas. Existe todavía una dificultad, por ejemplo, en el hecho de que sólo se pueden tomar raíces cuadradas de los números positivos (o cero) y no de los negativos. Desde el punto de vista matemático —y dejando de lado, por el momento, cualquier cuestión acerca de la conexión

* Recuérdese que la notación "10^{20}" representa el número 100000000000000000000, en donde el 1 esta seguido de veinte 0s.

directa con el mundo físico— resulta extraordinariamente conveniente poder extraer raíces cuadradas de números negativos tanto como de números positivos. Postulemos simplemente, o "inventemos", una raíz cuadrada para el número -1. La denotaremos por el símbolo "i", de modo que tenemos

$$i^2 = -1$$

La cantidad i no puede ser, por supuesto, un número real puesto que el producto de un número real por sí mismo es siempre positivo (o cero, si el propio número es el cero). Por esta razón se ha aplicado convencionalmente el término *imaginario a* los números cuyos cuadrados son negativos. Sin embargo, es importante resaltar el hecho de que estos números "imaginarios" no son menos reales que los números "reales" a los que estamos acostumbrados. Como he señalado antes, la relación entre tales números "reales" y la realidad *física* no es tan directa o irresistible como pudiera parecer a primera vista, al implicar, como lo hace, una idealización matemática de resolución infinita para la que no hay *a priori* una clara justificación en la naturaleza.

Teniendo una raíz cuadrada para -1 no supone ahora gran esfuerzo proporcionar raíces cuadradas para *todos* los números reales. En efecto, si *a* es un número real positivo, entonces la cantidad

$$i \times \sqrt{a}$$

es una raíz cuadrada del número real negativo $-a$. (Hay también otra raíz cuadrada, a saber, $-i \times \sqrt{a}$.) ¿Qué sucede con el propio i? ¿Tiene una raíz cuadrada? Seguro que la tiene. Se comprueba fácilmente que la cantidad

$$\frac{1+i}{\sqrt{2}}$$

(y también el negativo de esa cantidad) elevada al cuadrado da i. ¿Tiene *este* número una raíz cuadrada? De nuevo la respuesta es sí; el cuadrado de

$$\sqrt{\frac{1+1/\sqrt{2}}{2}} + i\sqrt{\frac{1-1/\sqrt{2}}{2}}$$

o de su negativo es en efecto $(1 + i)/\sqrt{2}$.

Nótese que al formar tales cantidades nos hemos permitido sumar números reales con números imaginarios, tanto como multiplicar nuestros números por números reales arbitrarios (o dividirlos por números reales diferentes de cero, que es lo mismo que multiplicar por sus recíprocos). Los objetos resultantes son los llamados *números complejos*. Un número complejo es un número de la forma

$$a + ib,$$

donde a y b son números reales llamados *parte real y parte imaginaria*, respectivamente, del número complejo. Las reglas para sumar y multiplicar tales números se siguen de las reglas ordinarias del álgebra, con la regla añadida de que $i^2 = -1$:

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

$$(a + ib) \times (c + id) = (ac - bd) + i(ad - bc).$$

Ahora sucede algo notable. Nuestro motivo para introducir este sistema de números había sido proporcionar la posibilidad de que siempre se pudieran tomar raíces cuadradas. El sistema consigue este cometido, aunque esto no es en sí mismo todavía obvio. Pero consigue mucho más: pueden obtenerse sin problemas raíces cúbicas, raíces quintas, raíces noventainueve-avas, raíces de orden π, raíces de orden (1 + i), etc. (como fue capaz de demostrar el gran matemático del siglo XVIII Leonhard Euler). Como un ejemplo más de la magia de los números complejos examinemos las fórmulas trigonométricas de apariencia algo complicada que aprendimos en la escuela: los senos y cosenos de la suma de dos ángulos

$$\text{sen}\,(A + B) = \text{sen}\,A \cos B + \cos A\ \text{sen}\,B,$$

$$\cos\,(A + B) = \cos A \cos B - \text{sen}\,A\ \text{sen}\,B,$$

son simplemente las partes imaginaria y real respectivamente de la mucho más simple (¡y mucho más fácil de recordar!) ecuación compleja.*

$$e^{iA+iB} = e^{iA}e^{iB}$$

Todo lo que necesitamos saber aquí es la "fórmula de Euler" (aparentemente también obtenida muchos años antes que Euler por el famoso matemático inglés del siglo XVI Roger Cotes)

$$e^{iA} = \cos A + i\ \text{sen}\,A,$$

que substituimos en la expresión anterior. La expresión resultante es

$$\cos(A + B) + i\ \text{sen}(A + B) = (\cos A + i\ \text{sen}\ A)(\cos B + i\ \text{sen}\ B)$$

y llevando a cabo la multiplicación en el segundo miembro obtenemos las relaciones trigonométricas deseadas. Más aún, cualquier ecuación algebraica

$$a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \ldots + a_3 z^3 = 0$$

(en la que a_0 , a_1 ,..., a_n son números complejos, con $a_n \neq 0$) puede satisfacerse siempre para un cierto z. Por ejemplo, existe un número complejo z que satisface la ecuación

$$z^{102} + 999z^{33} - \pi z^2 = -417 + i$$

aunque esto no es obvio en modo alguno. El hecho general es denominado a veces "teorema fundamental del álgebra". Varios matemáticos del siglo XVIII lucharon por demostrar este resultado. Ni siquiera Euler encontró un argumento general satisfactorio. Más tarde, en 1831, el gran matemático y científico Carl Friedrich Gauss dio una línea de argumentación sorprendentemente original y obtuvo la primera demostración general. Un ingrediente clave de esta demostración consistía en representar *geométricamente* los números complejos y, una vez hecho esto, utilizar un argumento topológico.#

---

* La cantidad e = 2.7182818285... (la base de los logaritmos naturales, y un número irracional de una importancia en matemáticas comparable a la de π) viene definida por

$$e = 1 + 1/1 + 1/(1x2) + 1/(1x2x3) + \ldots,$$

y e^z significa la z-ésima potencia de e, donde tenemos, como resultado

$$e^z = 1 + z/1 + z^2/(1x2) + z^3/(1x2x3) + \ldots,$$

La palabra "topológico" se refiere al tipo de geometría — denominada a veces "geometría de la lámina elástica" — en la que no se consideran las distancias reales, y sólo tienen importancia las propiedades de continuidad de los objetos.

En realidad, Gauss no fue el primero en utilizar una descripción geométrica de los números complejos. Wallis lo había hecho, de forma más tosca, unos doscientos años antes, aunque no la había utilizado con un efecto tan poderoso como lo hizo Gauss. El nombre que va asociado normalmente a esta representación geométrica de los números complejos es el de Jean Robert Argand, un contable suizo, que la describió en 1806, aunque el topógrafo noruego Caspar Wessel había dado ya una descripción muy completa nueve años antes. De acuerdo con esta terminología convencional (aunque no totalmente exacta históricamente) me referiré a la representación geométrica estándar de los números complejos como el *plano de Argand.*

El plano de Argand es un plano euclídeo ordinario con coordenadas cartesianas estándar x e y, donde x indica la distancia horizontal (positiva hacia la derecha y negativa hacia la izquierda) y donde y indica la distancia vertical (positiva hacia arriba y negativa hacia abajo). El número complejo

$$z = x + iy$$

viene representado entonces por el punto del plano de Argand cuyas coordenadas son

$$(x, y)$$

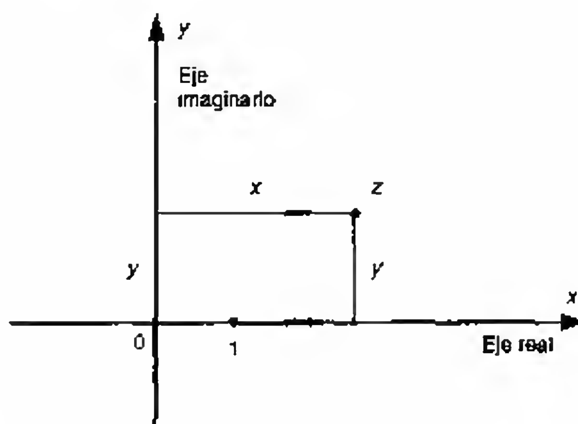


FIGURA III.8. *Un número complejo* z = x +iy *representado en el plano de Argand.*

Nótese que 0 (considerado como un número complejo) viene representado por el origen de coordenadas, y 1 viene representado por un punto en el eje x.

El plano de Argand proporciona simplemente un modo de organizar nuestra familia de números complejos en una imagen geométricamente útil. Cosas de este tipo no son realmente nuevas para nosotros. Ya estamos familiarizados con el modo en que se pueden organizar los números *reales* en una imagen geométrica: la imagen de una línea recta que se extiende indefinidamente en ambas direcciones. Un punto particular de la línea se etiqueta como 0 y otro punto se etiqueta como 1. El punto 2 está colocado de modo que su desplazamiento respecto a 1 es el mismo que el desplazamiento de 1 respecto a 0; el punto 1/2 el punto medio entre 0 y 1; el punto -1 está situado de modo que 0 esté en el punto medio entre -1 y 1, etc. El conjunto de los números reales representado de esta forma se conoce como la *recta real.* Para los números complejos tenemos que utilizar *dos* números reales como coordenadas, a saber a y b, para el número complejo $a + ib$.

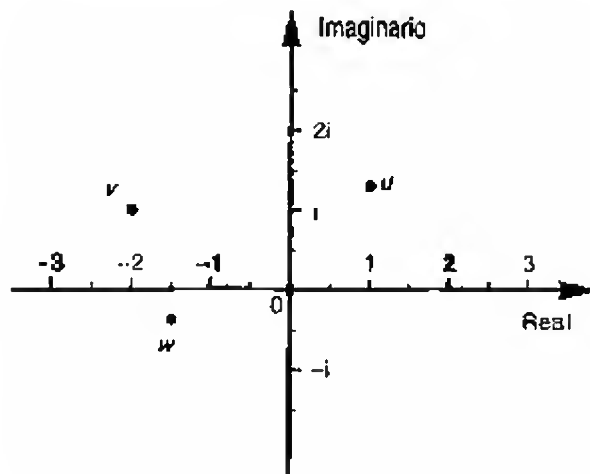


FIGURA III.9. *Localizaciones en el plano de Argand de* u=1+*i*1.3, v=-2+*i* *y* w=-1.5-*i*0.4

Estos dos números nos dan las coordenadas de un punto en un plano: el plano de Argand. Como ejemplo, he indicado en la fig. III.9 dónde estarían situados aproximadamente los números complejos

$$u=1+i1.3, \qquad v=-2+i, \qquad w=-1.5-i0.4,$$

Las operaciones algebraicas básicas de la suma y multiplicación de números complejos encuentran ahora una forma geométrica clara. Consideremos primero la suma. Supongamos que *u y v* son dos números complejos representados en el plano de Argand de acuerdo con el esquema anterior. Entonces su suma *u + v* viene representada como la suma vectorial" de los dos puntos; es decir, el punto *u + v* está en el lugar que completa el paralelogramo formado por w, v y el origen 0. No es muy difícil ver que esta construcción da efectivamente la suma (véase fig. III.10), pero omito aquí el argumento.

El producto *uv* tiene también una interpretación geométrica clara (ver fig. III. 11), que es quizá un poco más difícil de ver. (De nuevo omito el argumento.) El ángulo subtendido en el origen entre 1 y *uv* es la suma de

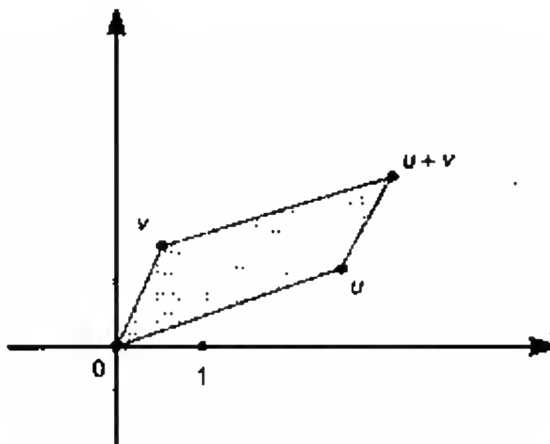


FIGURA III.10. *La suma* u + v *de dos números complejos* u *y* v *se obtiene mediante la ley del paralelogramo*

los ángulos entre 1 y u y entre 1 y v (todos los ángulos se miden en sentido contrario a las agujas del reloj), y la distancia desde el origen a uv es el producto de las distancias desde el origen a u y a v. Esto equivale a decir que el triángulo formado por 0, v y uv es semejante (y orientado de la misma forma) al triángulo formado por 0, 1 y u. (El lector animoso que no esté familiarizado con estas construcciones puede ocuparse en comprobar que se siguen directamente de las reglas algebraicas para la suma

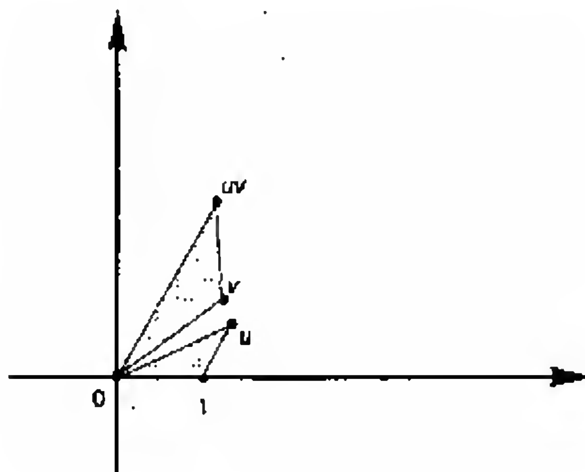


FIGURA III. 11. *El producto* uv *de dos números complejos* u *y* v *es tal que el triángulo formado por 0,* v *y* uv *es semejante al formado por 0, 1 y* u. *O lo que es equivalente*: *la distancia desde 0 a* uv *es el producto de las distancias desde 0 a* u *y a* v, *y el ángulo que forma* uv *con el eje real* (*horizontal*) *es la suma de los ángulos que forman* u *y* v *con dicho eje.*

y la multiplicación de números complejos que se dieron antes, junto con las identidades trigonométricas arriba expresadas.)

CONSTRUCCIÓN DEL CONJUNTO DE MANDELBROT

Estamos ahora en disposición de ver cómo se define el conjunto de Mandelbrot. Sea z algún número complejo escogido arbitrariamente. Cualquiera que sea este número complejo estará representado por algún punto en el plano de Argand. Consideremos ahora la *aplicación* según la cual z es reemplazado por un *nuevo* número complejo dado por

$$z \to z^2 + c,$$

donde c es otro número complejo *fijo* (es decir, dado). El número $z^2 + c$ estará representado por algún nuevo punto del plano de Argand. Por ejemplo, si se da c como 1.63 - i 4.2, entonces z se aplicará según

$$z \to z^2 + 1.63 - i4.2$$

de modo que, en particular, 3 será reemplazado por

$$3^2 + 1.63 - i4.2 = 9 + 1.63 - i4.2 = 10.63 - i4.2$$

y -2.7 + i0.3 será reemplazado por

$$(-2.7 + i0.3)^2 + 1.63 - i4.2$$

$= (-2.7)^2 - (0.3)^2 + 1.63 + i\{2(-2.7)(0.5) - 4.2\}$

$= 8.83 - i5.82.$

Cuando tales números se complican es mejor realizar los cálculos con una computadora.

Ahora, cualquiera que sea *c*, el número particular 0 se reemplaza, según este esquema, por el número dado *c*. ¿Qué ocurre con el propio *el* Éste debe ser reemplazado por el número $c^2 + c$. Supongamos que continuamos este proceso y aplicamos el reemplazamiento al número $c^2 + c$; entonces obtenemos

$$(c^2+c)^2 + c = c^4 + 2c^3 + c^2 + c.$$

Iteremos de nuevo el reemplazamiento, aplicándolo ahora al número anterior para obtener

$$(c^4 + 2c^3 + c^2 + c)^2 + c = c^8 + 4c^7 + 6c^6 + 6c^5 + 5c^4 + 2c^3 + c^2 + c$$

y luego otra vez a este número y así sucesivamente. Obtenemos una secuencia de números complejos, empezando con 0:

$$0,\ c,\ c^2 + c,\ c^4 + 2c^3 + c^2 + c, \ldots$$

Ahora bien, si hacemos esto con *ciertas* elecciones del número complejo *c*, la secuencia de números que obtenemos así nunca se aleja mucho del origen del plano de Argand; dicho con mayor exactitud, la secuencia permanece *acotada* para tales elecciones de *c*, lo que equivale a decir que todo número de la secuencia está dentro de algún *círculo fijo* centrado en el origen (véase fig. III. 12) Un buen ejemplo en donde ocurre esto es el caso c = 0, ya que en este caso todos los miembros de la secuencia resultan ser 0. Otro ejemplo de comportamiento acotado ocurre para c = -1, pues entonces la secuencia es: 0, -1, 0, -1, 0,-1,...; y un ejemplo más ocurre para c = i, siendo entonces la secuencia 0, i, i - 1, -i, i - 1, -i, i - 1, -i,... Sin embargo, para otros números complejos *c* la secuencia se aleja más y más hasta una distancia indefinida del origen; es decir, la secuencia está *no acotada*, *y* no puede estar contenida dentro de ningún círculo fijo. Un ejemplo de este último comportamiento ocurre para *c* = 1, pues entonces la secuencia es 0, 1, 2, 5, 26, 677, 458330,...; también sucede esto para *c* = -3, siendo en este caso la secuencia 0, -3, 6, 33, 1086,...; y también para *c* = i - 1, siendo la secuencia 0, i - 1, -i - 1, -1 + 3i,-9-i5, 55 + i91,-5257 + i10011,...

El *conjunto de Mandelbrot*, es decir, la región *en negro* de nuestro mundo de Tor'Bled-Nam, es precisamente la región del plano de Argand que consta de los puntos *c* para los que la secuencia permanece acotada. La región *blanca* consta de los puntos *c* para los que la secuencia es no acotada. Las imágenes detalladas que vimos antes estaban dibujadas a partir de datos de salida de computadoras. La computadora recorrerá sistemáticamente las posibles elecciones del número complejo c; para cada elección de c calculará la secuencia 0, *c*, $c^2 + c$,...y decidirá, según algún criterio apropiado, si la secuencia permanece acotada o no. Si *está* acotada, la computadora dispone que un punto negro aparezca en la pantalla en el punto correspondiente a *c*. Si no está acotada, la computadora dispone que aparezca un punto blanco. En definitiva, para cada punto de la pantalla en el rango considerado el ordenador decidirá si el punto debe estar coloreado blanco o negro.

La complejidad del conjunto de Mandelbrot es muy notable, sobre todo a la vista de que la definición de este conjunto es, como definición

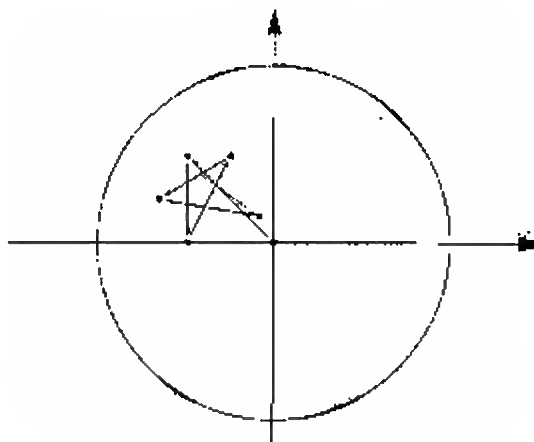


FIGURA III. 12. *Una secuencia de puntos en el plano de Argand está* acotada *si existe algún círculo fijo que contiene todos los puntos.* (*Esta iteración particular comienza con cero y tiene* c= -1/2 + 1/2 *i*.)

matemática, sorprendentemente simple. Sucede también que la estructura general de este conjunto no es muy sensible a la forma algebraica exacta de la aplicación $z \rightarrow z^2 + c$ que hemos escogido. Muchas otras aplicaciones iteradas complejas (*v.g.* $z \rightarrow z^3 + iz^2 + c$) darán estructuras extraordinariamente similares (siempre que escojamos un número apropiado para empezar, quizá no 0, sino un número cuyo valor esté caracterizado por una regla matemática clara para cada elección apropiada de la aplicación). Existe, en efecto, una especie de carácter absoluto o universal para estas estructuras "de Mandelbrot", con respecto a aplicaciones iteradas complejas. El estudio de estas estructuras constituye por sí mismo una rama de las matemáticas conocida como *sistemas dinámicos complejos*.

¿REALIDAD PLATÓNICA DE LOS CONCEPTOS MATEMÁTICOS?

Hasta qué punto son "reales" los objetos del mundo del matemático?. Desde un cierto punto de vista parece que no puede haber nada real en ellos. Los objetos matemáticos son sólo conceptos; son idealizaciones mentales que hacen los matemáticos, a menudo estimulados por el orden aparente de ciertos aspectos del mundo que nos rodea, pero idealizaciones mentales en cualquier caso. ¿Pueden ser algo más que meras construcciones arbitrarias de la mente humana? Al mismo tiempo parece que existe alguna realidad profunda en estos conceptos matemáticos que va más allá de las elucubraciones mentales de un matemático particular. En lugar de ello, es como si el pensamiento matemático estuviese siendo guiado hacia alguna verdad exterior —una verdad que tiene realidad por sí misma y que sólo se nos revela parcialmente a alguno de nosotros.

El conjunto de Mandelbrot proporciona un ejemplo sorprendente. Su estructura maravillosamente elaborada no fue la invención de ninguna persona, ni el diseño de un equipo de matemáticos. El propio Benoit Mandelbrot, el matemático polaco-estadounidense (protagonista de la teoría fractal) que primero[3] estudió el conjunto no tenía ninguna concepción previa acerca de la fantástica elaboración inherente al mismo, aunque sabía que estaba en la pista de algo muy

[3] Realmente existe cierta controversia acerca de quién fue el primero que dio con este conjunto (véase Brooks y Matelski, 1981, Mandelbrot, 1989); pero justamente el hecho de que pueda *haber* tal disputa confiere mayor crédito a la idea de que el hallazgo del conjunto fue algo más parecido a un descubrimiento que a una invención.

interesante. De hecho, cuando empezaron a surgir sus primeras imágenes de computadora, él tuvo la impresión de que las estructuras difusas que estaba viendo eran el resultado de un mal funcionamiento de la computadora (Mandelbrot, 1986).

Sólo más tarde llegó a convencerse de que estaban realmente en el propio conjunto. Además, los detalles completos de la compleja estructura del conjunto de Mandelbrot no pueden ser aprehendidos realmente por ninguno de nosotros, ni pueden ser completamente revelados por una computadora. Parecería que esta estructura no es sólo parte de nuestras mentes sino que tiene una realidad autónoma. Cualquiera que sea el entusiasta matemático o computadora que decida examinar el conjunto, encontrará aproximaciones a la *misma* estructura matemática fundamental. No hay ninguna verdadera diferencia que dependa de la computadora que se utilice para hacer los cálculos (siempre que la computadora tenga una precisión suficiente), aparte del hecho de que las diferencias en velocidad y memoria de la computadora, y capacidades de representación gráfica, puedan conducir a diferencias en los detalles finos que saldrán a la luz y en la velocidad con que se produce este detalle. La computadora está siendo utilizado esencialmente de la misma forma en que el físico experimental utiliza un aparato experimental para explorar la estructura del mundo físico. El conjunto de Mandelbrot no es una invención de la mente humana; fue un descubrimiento. Al igual que el Monte Everest, el conjunto de Mandelbrot está *ahí.*

De modo análogo, el propio sistema de los números complejos tiene una realidad profunda e intemporal que va bastante más allá de las construcciones mentales de cualquier matemático particular. Los comienzos de una apreciación de los números complejos procedían de la obra de Girolamo Cardano. Este era un italiano que vivió entre 1501-1576, médico de formación, jugador y confeccionador de horóscopos (en cierta ocasión hizo un horóscopo de Cristo), y que escribió un importante e influyente tratado de álgebra, el *Ars Magna* en 1545. En éste desarrolló la primera expresión completa para la solución (en términos de radicales, esto es, raíces n-ésimas) de una ecuación cúbica general.* Él había notado, sin embargo, que en cierto tipo de casos —los llamados "irreducibles", en los que la ecuación tiene tres soluciones reales— se veía obligado a tomar, en cierto paso de su expresión, la *raíz cuadrada de un número negativo.* Aunque esto era un enigma para él, se dio cuenta de que si se permitía *tomar* esa raíz cuadrada, y *sólo* entonces, podía expresar la respuesta completa (siendo la respuesta final siempre real). Más tarde, en 1572, Raphael Bombelli, en una obra titulada *l'Algebra*, extendió el trabajo de Cardano y comenzó el estudio del actual álgebra de los números complejos.

Aunque al principio puede parecer que la introducción de tales raíces cuadradas de números negativos es sólo un artificio —una invención matemática diseñada para conseguir un determinado propósito— se hizo claro más adelante que estos objetos consiguen mucho más que aquello para lo que fueron diseñados originalmente. Como mencioné arriba, aunque el objetivo original de la introducción de los números complejos era poder tomar raíces cuadradas sin problemas, al introducir tales números nos encontramos, como premio añadido, con la potencialidad de tomar cualquier otro tipo de raíz o resolver cualquier ecuación algebraica. Más adelante encontraremos muchas otras propiedades mágicas que poseen estos números complejos, propiedades que no sospechábamos al principio. Estas propiedades están *ahí.* No fueron puestas por Cardano, ni por Bombelli, ni Wallis, ni Coates, ni Euler, ni Wessel, ni Gauss, a pesar de la indudable perspicacia de éstos y otros grandes matemáticos; dicha magia era inherente a la propia estructura que estaban descubriendo gradualmente. Cuando Cardano introdujo sus

* Basado en parte en la obra previa de Scipione del Ferro y de Tartaglia.

números complejos no podía sospechar las muchas propiedades mágicas que se iban a seguir de ello —propiedades que aparecen con nombres diversos, como la fórmula integral de Cauchy, el teorema de la aplicación de Riemann, la propiedad de extensión de Lewy—. Éstas, y muchos otros hechos notables, son propiedades de los mismos números, sin ninguna modificación adicional, que Cardano encontró por primera vez alrededor de 1539.

¿Es la matemática invención o descubrimiento? Cuando los matemáticos obtienen sus resultados ¿están produciendo solamente elaboradas construcciones mentales que no tienen auténtica realidad, pero cuyo poder y elegancia basta simplemente para engañar incluso a sus inventores haciéndoles creer que estas construcciones mentales son "reales"? ¿O están descubriendo realmente verdades que estaban ya "ahí", verdades cuya existencia es independiente de las actividades de los matemáticos? Creo que, por ahora, debe quedar muy claro para el lector que me adhiero al segundo punto de vista, más que al primero, al menos con respecto a estructuras como los números complejos y el conjunto de Mandelbrot.

Pero quizá la cuestión no sea tan sencilla como esto. Como ya he dicho, hay cosas en las matemáticas, tales como los ejemplos que acabo de citar, para las que el término "descubrimiento" es mucho más apropiado que "invención". Existen los casos en que sale de la estructura mucho más de lo que se introdujo al principio. Podríamos adoptar el punto de vista de que en tales casos los matemáticos han tropezado con "obras de Dios". Sin embargo, existen otros casos en los que la estructura matemática no tiene esa compulsiva unicidad; por ejemplo, cuando en medio de la demostración de algún resultado el matemático encuentra necesario introducir alguna construcción artificial, y de ningún modo única, para conseguir algún fin muy específico. En tales casos, no es probable que se obtenga nada más de la construcción que lo que se puso al principio, y la palabra "invención" parece más apropiada que "descubrimiento". Estas son "obras del hombre". Desde este punto de vista, los auténticos descubrimientos matemáticos serían considerados, de forma general, como consecuciones o aspiraciones más altas que lo que serían las "meras" invenciones.

Tales clasificaciones no son muy diferentes de las que podríamos utilizar en las artes o la ingeniería. Las grandes obras de arte están "más cerca de Dios" que las obras menores. Es un sentir no poco común entre los mayores artistas que en sus obras están revelando verdades eternas que tienen algún tipo de existencia etérea previa,* mientras que sus obras menores podrían ser más arbitrarias, de la misma naturaleza de las meras construcciones mortales. De modo análogo, una innovación de bella sencillez en ingeniería, con la que abre una enorme perspectiva para la aplicación de alguna idea simple e inesperada, podría ser descrita con propiedad como un descubrimiento más que una invención.

No obstante, después de hacer estos comentarios no puedo evitar el sentimiento de que, en el caso de las matemáticas, la creencia en algún tipo de existencia etérea y eterna, al menos para los conceptos más profundamente matemáticos, es mucho más fuerte que en los otros casos. *Hay* en tales ideas matemáticas una compulsiva unicidad y universalidad que parecen ser de un orden diferente del que se pudiera esperar en las artes o la ingeniería. El punto de vista de que los conceptos matemáticos podrían existir en ese sentido etéreo e intemporal fue planteado en tiempos antiguos (c. 360 a.C.) por el gran filósofo griego Platón. En consecuencia, este punto de

* Como dijo el famoso escritor argentino Jorge Luis Borges: "...un buen poeta es un descubridor antes que un inventor...".

vista es calificado a veces de platonismo matemático. Tendrá gran importancia para nosotros más tarde.

En el capítulo I discutí con cierta amplitud el punto de vista de la IA *fuerte*, que supone que los fenómenos mentales encuentran su existencia dentro de la idea matemática de un algoritmo. En el capítulo II subrayé el punto de que el concepto de un algoritmo es una idea profunda e "infusa". En ese capítulo he argumentado que tales ideas matemáticas "infusas" tendrían algún tipo de existencia intemporal, independientes de las nuestras terrenales. ¿No da este punto de vista algún crédito al punto de vista de la IA fuerte, al proporcionar la posibilidad de un tipo de existencia etérea para los fenómenos mentales? Es perfectamente razonable —y aún especularé más tarde en favor de un punto de vista no muy diferente de éste; pero si los fenómenos mentales pueden encontrar un hogar de este tipo general, no creo que pueda ser en el concepto de un algoritmo. Sería necesario algo mucho más sutil. El hecho de que las cosas algorítmicas constituyen una parte muy estrecha y limitada de las matemáticas será un aspecto importante de las discusiones siguientes. Empezaremos a ver algo del alcance y sutileza de las matemáticas no algorítmicas en el próximo capítulo.

IV. VERDAD, DEMOSTRACIÓN E INTUICIÓN DIRECTA

EL PROGRAMA DE HILBERT PARA LAS MATEMÁTICAS

¿QUE ES LA VERDAD? ¿Cómo formamos nuestros juicios acerca de lo que es verdadero y lo que es falso? ¿Estamos siguiendo simplemente un *algoritmo*, sin duda favorecido sobre otros menos efectivos por el poderoso proceso de selección natural? ¿O puede existir algún otro camino, posiblemente no algorítmico —tal vez perspicacia, instinto o intuición directa— para descubrir la verdad? Esta parece una pregunta difícil. Nuestros juicios se basan en combinaciones complejas e interconectadas de datos sensoriales, razonamientos y conjeturas. Además, en muchas situaciones ordinarias los criterios sobre lo que realmente *es* verdadero y lo que es falso pueden variar. Para simplificar la pregunta, consideremos sólo la verdad *matemática*. ¿Cómo formamos nuestros juicios —quizá incluso nuestro conocimiento "seguro" respecto a las cuestiones matemáticas? Aquí, por lo menos, las cosas deberían ser más nítidas. No debería haber ningún problema sobre lo que realmente es verdadero y lo que realmente es falso, ¿o debería haberlo? ¿Qué *es*, realmente, la verdad matemática? Esta es una cuestión muy vieja, que se remonta al menos a la época de los primeros filósofos y matemáticos griegos —y, sin duda, aún más atrás. No obstante, en los últimos cien años, más o menos, se han despejado mucho las cosas y se han conseguido intuiciones *nuevas y* sorprendentes. Son estos nuevos desarrollos los que trataremos de comprender. Los resultados obtenidos son fundamentales y afectan a la cuestión misma de si nuestros procesos de pensamiento pueden ser de naturaleza completamente algorítmica. Es importante que entendamos esto. A finales del siglo XIX las matemáticas habían hecho grandes progresos, debido en parte al desarrollo de métodos de demostración cada vez más poderosos. (David Hilbert y Georg Cantor, a quienes ya hemos mencionado antes, y el gran matemático francés Henri Poincaré, a quien citaremos más adelante, estaban en la vanguardia de estos desarrollos.) En consecuencia, los matemáticos habían adquirido confianza en el uso de estos poderosos métodos. Muchos de estos métodos implicaban considerar conjuntos[*] con un infinito número de miembros, y las demostraciones tenían a menudo éxito precisamente porque era posible considerar tales conjuntos como objetos reales, unidades completas y existentes, y no simplemente con una existencia potencial. Muchas de estas ideas habían brotado del muy original concepto de *números infinitos o transfinitos* de Cantor, que éste había desarrollado coherentemente utilizando conjuntos infinitos. (Vimos algo de esto en el capítulo anterior.)

Sin embargo, esta confianza quedó hecha añicos cuando en 1902 el lógico y filósofo británico Bertrand Russell presentó su ahora célebre paradoja (ya anticipada por Cantor, y descendiente directa del argumento del "corte diagonal"). Para comprender el argumento de Russell necesitamos primero tener alguna noción de lo que está implícito en la consideración de los conjuntos como unidades completas. Podemos imaginar un conjunto que está caracterizado en términos de una *propiedad* concreta. Por ejemplo, el conjunto de las cosas *rojas* está caracterizado por la propiedad de ser *rojo*: una cosa pertenece a dicho conjunto si y sólo si es roja. Esto nos permite invertir el planteamiento y hablar de una propiedad en términos de un solo objeto, a saber, el conjunto de todas las cosas que tienen dicha propiedad. Con este punto de vista, "lo rojo" es el conjunto de todas las cosas rojas. (Podemos concebir también que algunos

[*] Un *conjunto* significa simplemente una colección de cosas —objetos físicos o conceptos matemáticos— que puede considerarse como un todo. En matemáticas, los elementos de un conjunto son muy a menudo conjuntos ellos mismos, ya que pueden reunirse conjuntos para formar otros conjuntos. En consecuencia, se pueden considerar conjuntos de conjuntos, o conjuntos de conjuntos de conjuntos, etcétera.

otros conjuntos están "ahí", con sus elementos caracterizados por alguna propiedad no tan simple.)

Esta idea de definir los conceptos en términos de conjuntos era fundamental para el procedimiento, introducido en 1884 por el influyente lógico germano Gottlob Frege, según el cual los *números* pueden definirse en términos de conjuntos. Por ejemplo, ¿qué queremos decir mediante el número 3? Sabemos que es la propiedad de "treidad" pero ¿qué es el 3 en sí mismo? Ahora bien, la "treidad" es una propiedad de *colecciones* de objetos, es decir, es una propiedad de *conjuntos*: un conjunto tiene la propiedad particular de "treidad" si y sólo si el conjunto tiene exactamente tres miembros. Por ejemplo, el conjunto de los ganadores de medallas en una competición olímpica concreta tiene esta propiedad de "treidad". También la tienen el conjunto de los neumáticos en un triciclo, o el conjunto de las hojas de un trébol normal, o el conjunto de las soluciones de la ecuación $x^3 - 6x^2 + 11x - 6 = 0$. ¿Cuál es, entonces, la definición de Frege del número 3? Según Frege, 3 debe ser un conjunto de conjuntos: el conjunto de *todos* los conjuntos con la propiedad de "treidad"[1]. De este modo, un conjunto tiene tres miembros si y sólo si pertenece al conjunto 3 de Frege.

Esto puede parecer una definición circular pero, en realidad, no lo es. En general, podemos definir los *números* como totalidades de conjuntos equivalentes, donde "equivalentes" significa en este caso "que tienen elementos que pueden ser emparejados uno-a-uno" (es decir, en términos ordinarios, "que tienen el mismo número de elementos"). El número 3 es entonces aquel conjunto particular tal que uno de sus miembros es un conjunto que contiene, por ejemplo, simplemente una manzana, una naranja y una pera. Nótese que esta es una definición de "3" bastante diferente de la de 3 de Church dada. Pueden darse también otras definiciones que son bastante más populares en estos días.

Ahora bien, ¿qué hay de la paradoja de Russell? Esta se refiere a un conjunto *R* definido de la manera siguiente:

R es el conjunto de todos los conjuntos que no son miembros de sí mismos.

Por lo tanto, *R* es ciertamente una colección de conjuntos; y el criterio para que un conjunto X pertenezca a esta colección es que el mismo conjunto X no se encuentre entre sus *propios* elementos.

¿Es absurdo suponer que un conjunto pueda ser elemento de sí mismo? Ciertamente no. Consideremos, por ejemplo, el conjunto *I* de los conjuntos *infinitos* (conjuntos con infinitos elementos). Ciertamente hay una infinidad de conjuntos infinitos *diferentes*, de modo que *I* es él mismo infinito. Por consiguiente, *I* pertenece realmente a sí mismo. Entonces ¿dónde está la paradoja de Russell? Preguntemos: ¿es el propio conjunto *R* de Russell un elemento de sí mismo o no lo es? Si *no* es un miembro de sí mismo entonces debería pertenecer a *R* puesto que *R* consta precisamente de aquellos conjuntos que no son elementos de sí mismos. Por consiguiente,

[1] Al considerar conjuntos cuyos miembros pueden ser a su vez conjuntos debemos tener cuidado en distinguir entre los miembros de ese conjunto y los miembros de los *miembros* de dicho conjunto. Por ejemplo, supongamos que S es el conjunto de los *subconjuntos no vacíos* de otro conjunto *T*, en donde los miembros de *T* son una manzana y una naranja. *T* tiene la propiedad de *par* y no de "treidad", pero S tiene realmente la propiedad de "treidad"; en efecto, los tres miembros de S son: un conjunto que sólo contiene una manzana, un conjunto que sólo contiene una naranja, y un conjunto que contiene una manzana y una naranja, tres conjuntos en total, que son los *tres* miembros de S. Análogamente, el conjunto cuyo único miembro es el *conjunto vacío* posee la propiedad de unidad no de nulidad, tiene *un* miembro: el propio conjunto vacío. Por supuesto, el propio conjunto vacío tiene cero miembros

R pertenece a *R*, después de todo: una contradicción. Por otro lado, si *R* es un elemento de sí mismo, entonces, puesto que el mismo es realmente *R*, pertenece al conjunto cuyos miembros se caracterizan por *no* ser miembros de sí mismos, es decir, no es miembro de sí mismo después de todo: de nuevo una contradicción.*

Esta consideración no es una frivolidad. Russell estaba simplemente utilizando, de una forma bastante extrema, el mismo tipo de razonamiento matemático general de teoría de conjuntos que los matemáticos estaban empezando a emplear en sus demostraciones. Evidentemente las cosas se les habían ido de las manos y resultaba conveniente ser mucho más preciso sobre el tipo de razonamiento que utilizarían en adelante. Era obviamente necesario que los razonamientos permitidos estuvieran libres de contradicción y permitieran, a partir de enunciados que se sabe previamente son verdaderos, sólo derivar también enunciados verdaderos. El propio Russell, junto con su colega Alfred North Whitehead, llegó a desarrollar un sistema matemático de axiomas y reglas de inferencia altamente formalizado, cuyo propósito era lograr que fuera posible traducir en su esquema todos los tipos de razonamientos matemáticos correctos. Las reglas estaban cuidadosamente seleccionadas para impedir los razonamientos que conducían a la paradoja de Russell. El esquema concreto que presentaron Russell y Whitehead constituyó una obra monumental. Sin embargo, era algo incómodo y los tipos de razonamiento matemático que incorporaba resultaban ser bastante limitados. El gran matemático David Hilbert, a quien mencionamos en el capítulo II, se embarcó en la tarea de establecer un esquema mucho más manejable y comprensible, que incluía *todos* los tipos de razonamiento matemáticamente correctos para cualquier área matemática particular. Además, Hilbert pretendía que era posible *demostrar* que el esquema estaba libre de contradicción. Sólo entonces las matemáticas estarían situadas, de una vez por todas, sobre unos fundamentos inobjetables.

Sin embargo, las esperanzas de Hilbert y sus seguidores quedaron defraudadas cuando, en 1931, el brillante lógico matemático austríaco de 25 años, Kurt Gödel, presentó un sorprendente teorema que destrozaba el programa de Hilbert. Lo que Gödel demostró era que *cualquiera* de estos sistemas matemáticos precisos ("formales") de axiomas y reglas de inferencia, siempre que sea lo bastante amplio como para contener descripciones de proposiciones aritméticas simples (como el "último teorema de Fermat" considerado en el capítulo II) y siempre que esté libre de contradicción, debe contener algunos enunciados que no son demostrables ni indemostrables con los medios permitidos dentro del sistema. La verdad de tales enunciados es así *indecidible* mediante los procedimientos aceptados. De hecho, Gödel pudo demostrar que el mismo enunciado de la consistencia del propio sistema axiomático —cuando se codifica en forma de una proposición aritmética adecuada— debe ser una de estas proposiciones *indecidibles*. Será importante que comprendamos la naturaleza de esta *indecidibilidad*. Veremos por qué el argumento de Gödel rompe el núcleo mismo del programa de Hilbert. Veremos también cómo el argumento de Gödel nos da la posibilidad, mediante intuición directa, de ir más allá de las limitaciones de cualquier sistema matemático formalizado particular bajo consideración. Comprender esto será crucial para gran parte de la discusión posterior.

* Existe una divertida manera de expresar la paradoja de Russell en términos ordinarios. Imaginemos una biblioteca en la que hay dos catálogos, en uno de los cuales se incluyen todos los libros de la biblioteca que hacen referencia a sí mismos, y en el otro se incluyen precisamente todos los libros que no hacen mención de sí mismos. ¿En cuál de estos dos debe ser incluido el segundo catálogo?

SISTEMAS MATEMÁTICOS FORMALES

Será preciso que seamos un poco más explícitos sobre lo que entendemos por un "sistema matemático formal de axiomas y reglas de inferencia". Debemos suponer que existe algún alfabeto de símbolos en cuyos términos se expresan nuestros enunciados matemáticos. Ciertamente estos símbolos deben ser adecuados para permitir una notación para los números naturales, de modo que la "aritmética" pueda incorporarse dentro de nuestro sistema. Podemos, si así lo queremos, utilizar simplemente la notación arábiga usual *0*, *1*, *2*, *3*,..., *9*, *10*, *11*, *12*,... para los números, aunque esto hace que la especificación de las reglas sea un poco más complicada de lo necesario. Resultará más simple si utilizamos, por ejemplo, 0, 01, 011, 0111, 01111,... para denotar la serie de los números naturales (o, a modo de compromiso, podríamos utilizar la notación binaria). No obstante, puesto que ello provocaría confusión en la discusión que sigue, me atendré a la notación arábiga usual en mis descripciones, cualquiera que sea la notación que el sistema pudiera usar *realmente*. Podríamos necesitar un símbolo "espacio" para separar las distintas "palabras" o "números" de nuestro sistema pero, ya que esto también puede resultar confuso, utilizaremos sólo una coma (,) para este propósito cuando sea necesario. También necesitaremos letras que denoten números naturales arbitrarios ("variables") — enteros o racionales — , aunque nos limitaremos aquí a los números naturales — , digamos *t*, *u*, *v*, *w*, *x*, *y*, *z*, *t*', *t*", *t*'", . . . Pueden ser necesarias letras con prima (*t*', *t*",...), ya que no queremos poner ningún límite definido al número de variables que pueden figurar en una expresión. Consideramos *la prima* (') como otro símbolo más dentro del sistema formal, de modo que el número real de *símbolos* permanece finito. Necesitaremos símbolos para las operaciones aritméticas básicas =, +, x, etc., quizá para varios tipos de paréntesis (,), [,], y para los símbolos *lógicos* tales como & ("y"), ⇒ ("implica"), V ("o"), ⇔ ("si y sólo si"), ~ ("no" o "no es el caso que..."). Además, nos harán falta los "cuantificadores" lógicos : *el cuantificador existencial* ∃ ("existe... tal que") y el *cuantificador universal* ∀ ("para todo ... tenemos"). Entonces podemos formar enunciados tales como el "último teorema de Fermat".

$$\sim\exists wxyz v[(x+1)^{w+3}+(y+1)^{w+3}+(z+1)^{w+3}]$$

(véase el capítulo II). (Podría haber escrito 0111 en lugar de 3, y para elevar a una potencia, utilizar una notación que se ajuste mejor al formalismo; pero, como ya he dicho, me estoy ateniendo a los símbolos tradicionales para no introducir confusiones innecesarias.) El enunciado anterior se lee (terminando en el primer paréntesis cuadrado):

"No es el caso que existan números naturales *w*, *x*, *y*, *z* tales que...",

Podemos también reescribir el último teorema de Fermat utilizando ∀:

$$\forall\, w,x,y,z\ [\sim (x+1)^{w+3}+(y+1)^{w+3}=(z+1)^{w+3}],$$

que se lee (hasta el símbolo "no" tras el primer paréntesis):

"Para todos los números naturales *w*, *x*, *y* , *z* no es el caso que...",

que es lógicamente equivalente a lo anterior.

También necesitamos letras para denotar proposiciones completas, y para ello utilizaré letras mayúsculas: *P*, *Q*, *R*, *S*,... Una proposición tal podría ser de hecho la anterior afirmación de Fermat:

$$F = \sim\exists wxyz\mathrm{v}[(x+1)^{w+3}+ (y+1)^{w+3} + (z+1)^{w+3}]$$

Una proposición puede también *depender* de una o más variables; por ejemplo, podríamos estar interesados en la afirmación de Fermat para alguna *potencia particular*[*] $w + 3$:

$$G(w) = \sim\exists wxyz\mathrm{v}[(x+1)^{w+3}+ (y+1)^{w+3} + (z+1)^{w+3}]$$

de modo que $G(0)$ afirma que "ningún cubo puede ser suma de cubos positivos", $G(1)$ afirma lo mismo para la cuarta potencia, y así sucesivamente. (Nótese que ahora falta la "w" tras "$\exists$".) La afirmación de Fermat es ahora que G(w) es válida para *todo w*:

$$F = \forall w\ [G(w)].$$

G( ) es un ejemplo de lo que se conoce como una *función proposicional*; es decir, una proposición que depende de una o más variables.

Los *axiomas* del sistema constituirán una lista finita de proposiciones generales cuya verdad, dados los significados de los símbolos, se supone evidente por sí misma. Por ejemplo, para proposiciones o funciones proposicionales arbitrarias P, Q, $R(\ )$, tendremos entre nuestros axiomas:

$$(P\&Q) \Rightarrow P,$$

$$\sim(\sim P) \Leftrightarrow P,$$

$$\exists x\ [R(x)] \Leftrightarrow \forall x[\sim R(x)],$$

cuya "verdad evidente" se comprueba inmediatamente a partir de sus *significados*. (El primero afirma simplemente que: "si P y Q son ambas verdaderas, entonces P es verdadera"; la segunda afirma la equivalencia entre "no es cierto que P es falsa" y "P es verdadera"; la tercera está ejemplificada por la equivalencia lógica de las dos formas de enunciar el último teorema de Fermat dadas más arriba.) Podríamos incluir también axiomas aritméticos básicos, tales como

$$\forall x,y\ [x + y = y + x]$$

$$\forall x,y,z\ [(x + y) \times z = (x \times z)+(y \times z)]$$

aunque podría ser preferible construir estas operaciones matemáticas a partir de algo más elemental y deducir estos enunciados como teoremas. Las *reglas de inferencia* serán cosas (evidentes) como:

a partir de P y P $\Rightarrow$ Q podemos deducir Q

a partir de $\forall x[R(x)]$ podemos deducir cualquier proposición que se obtenga sustituyendo x por un número natural concreto en $R(x)$.

Estas son instrucciones que nos dicen cómo podemos derivar nuevas proposiciones a partir de otras ya establecidas.

[*] Aunque todavía no se sabe si es verdadera la proposición general F de Fermat, se sabe que son verdaderas las proposiciones individuales $G(0)$, $G(1)$, $G(2)$, $G(3)$,.. .hasta aproximadamente $G(125\ 000)$. Es decir, se sabe que ningún cubo puede ser la suma de cubos positivos, ninguna potencia cuarta puede ser suma de potencias cuartas, etc., y así sucesivamente hasta el enunciado para la milésima potencia 125.

Ahora bien, partiendo de los axiomas y aplicando una y otra vez las reglas, podemos construir una larga lista de proposiciones. En cualquier etapa, podemos hacer entrar de nuevo en juego alguno de los axiomas, y seguir haciendo uso de las proposiciones que ya hemos añadido a nuestra creciente lista. Las proposiciones de cualquiera de estas listas se conocen como *teoremas* (aunque muchas de ellas sean bastante triviales o sin interés como enunciados matemáticos). Si tenemos una proposición concreta *P* que queremos *demostrar*, entonces trataremos de encontrar dicha lista, correctamente concatenada con arreglo a estas reglas, y que termina con nuestra proposición *P*. Tal lista nos proporcionaría una *demostración* de *P* dentro del sistema y, en consecuencia, *P* será un teorema.

La idea del programa de Hilbert consistía en encontrar, para cualquier área bien definida de las matemáticas, una lista de axiomas y reglas de inferencia suficientemente amplia que incorporara *todas* las formas de razonamiento matemático correcto apropiadas para dicha área. Tomemos la *aritmética* como nuestra área en cuestión de las matemáticas (donde se incluyen los cuantificadores ∃ y ∀, de modo que puedan formarse enunciados como el del último teorema de Fermat). No habría ninguna ventaja para nosotros en considerar aquí cualquier área matemática más general que ésta. La aritmética es *ya* bastante general para aplicar el método de Gödel. Según el programa de Hilbert, si se acepta que semejante sistema de axiomas y reglas de inferencia nos ha sido ya dado para la aritmética, entonces dispondremos de un criterio definido para la "corrección" de la demostración matemática de cualquier proposición aritmética. Existía la esperanza de que tal sistema de axiomas y reglas fuera *completo*, en el sentido de que nos permitiera en principio distinguir la verdad o falsedad de *cualquier* enunciado matemático que pueda formularse dentro del sistema.

Hilbert esperaba que para cualquier cadena de símbolos que represente una proposición matemática, digamos *P*, deberíamos ser capaces de demostrar o bien *P* o bien ~*P*, según sea *P* verdadera o falsa. Aquí debemos suponer que la cadena es *sintácticamente correcta* en su construcción, en donde "sintácticamente correcta" significa en esencia "gramaticalmente" correcta —es decir, que satisface todas las reglas gramaticales del formalismo, como que los paréntesis estén correctamente emparejados, etc.— de modo que *P* tiene un significado verdadero o falso bien definido. Si pudiera hacerse realidad la esperanza de Hilbert, esto nos dispensaría incluso de preocuparnos de lo que las proposiciones *significan*. *P* sería simplemente una cadena de símbolos sintácticamente correcta. A la cadena de símbolos *P* se le asignaría el valor de verdad **verdadero** si *P* es un teorema (es decir, si *P* es demostrable dentro del sistema), o se le asignaría el valor de verdad **falso** si, por el contrario, ~*P* es un teorema. Para que esto tenga sentido se requiere *consistencia* además de la *completitud*. Es decir, no deben existir cadenas de símbolos *P* para los que tanto *P* como ~*P* sean teoremas. De lo contrario, *P* podría ser **verdadera** y **falsa** al mismo tiempo.

La idea de que se puede prescindir de los significados de los enunciados matemáticos, considerándolos nada más como cadenas de símbolos en algún sistema matemático formal, es el punto de vista del *formalismo*. A ciertas personas les gusta esta idea, según la cual las matemáticas se convierten en una especie de "juego sin significado". Sin embargo, no es una idea que me seduzca. Es, en realidad, el "significado" —y no la ciega computación algorítmica— lo que constituye la substancia de las matemáticas. Afortunadamente, Gödel asestó un golpe devastador al formalismo. Veamos cómo lo hizo.

EL TEOREMA DE GÖDEL

Una parte del argumento de Gödel es muy complicada y realmente no es necesario que la examinemos. Pero la idea central era sencilla, bella y profunda; podremos apreciar esta idea. La parte complicada (que también contiene mucho ingenio) consistía en mostrar en detalle cómo se pueden codificar realmente las reglas de inferencia individuales del sistema formal, también el uso de sus diversos axiomas, en *operaciones aritméticas*. (Sin embargo, un aspecto de la parte profunda era hacer comprender que esto era algo fructífero.) Para llevar a cabo esta codificación necesitamos encontrar alguna forma conveniente de etiquetar las proposiciones mediante números naturales. Una forma consistiría simplemente en utilizar algún tipo de orden "alfabético" para las cadenas de símbolos del sistema formal con una longitud específica, dentro de una ordenación global según la longitud de la cadena. (Así, podrían ordenarse alfabéticamente las cadenas de longitud uno, seguidas de las cadenas de longitud dos, alfabéticamente ordenadas, seguidas de las cadenas de longitud tres, etc.) Esto se llama orden *lexicográfico*.* En realidad, Gödel utilizó originalmente un sistema de numeración más complicado, pero las diferencias no son importantes en este momento. Nos interesaremos especialmente en las *funciones proposicionales* que dependen de una *sola* variable, como es el caso de la $G(w)$ anterior. Sea la n-ésima de estas funciones proposicionales (en el orden elegido para las cadenas de símbolos) aplicada a w

Podemos permitirnos, si queremos, que nuestra numeración sea un poco "chapucera", de forma que algunas de estas expresiones no sean sintácticamente correctas. (Esto hace la codificación aritmética mucho más fácil que si tratáramos de omitir todas las expresiones sintácticamente incorrectas.) Si $P_n(w)$ es sintácticamente correcta será algún enunciado aritmético perfectamente bien definido que concierne a los dos números naturales n y w. Cuál sea exactamente este enunciado dependerá de los detalles del sistema de numeración específico que hayamos elegido. Eso pertenece a la parte complicada del argumento y no nos interesa aquí. Las cadenas de proposiciones que constituyen una *demostración* de algún teorema en el sistema pueden ser etiquetadas también mediante números naturales utilizando el esquema de ordenación escogido. Denotemos mediante

$$\Pi_n$$

la n-ésima demostración. (De nuevo, podemos utilizar una numeración "chapucera" en la que para algunos valores de n la expresión "Π_n" no es sintácticamente correcta y por lo tanto no demuestra ningún teorema.) Consideremos ahora la siguiente función proposicional, que depende del número natural w:

$$\sim\exists x\,[\Pi_x \text{ demuestra } P_w(w)]\ .$$

El enunciado dentro del paréntesis cuadrado se da parcialmente en palabras, pero es un enunciado perfecta y exactamente bien definido. Afirma que la x-ésima demostración es realmente una demostración de la proposición que constituye $P_w(\ )$ aplicada al propio valor w. El cuantificador existencial negado, fuera del paréntesis, sirve para eliminar una de las variables ("no existe un x tal que..."), de modo que nos queda una función proposicional aritmética que

* Podemos pensar en el orden lexicográfico como la ordenación normal de los números naturales escritos en "base $k + 1$ " utilizando, para los $k + 1$ dígitos, los diversos símbolos del sistema formal, junto con un nuevo "cero" que no se utiliza nunca. (Esta última complicación surge debido a que los números que comienzan por cero son los mismos que aquellos en los que se omite el cero.) Un sencillo orden lexicográfico de cadenas con nueve símbolos es el dado por los números naturales que pueden escribirse en notación decimal ordinaria sin el cero: 1, 2, 3,4,..., 8, 9, 11, 12,..., 21, 22 ..... 99, 111, 112,...

sólo depende de una variable w. La expresión global afirma que *no* existe demostración de $P_w(w)$. Supondré que está estructurada de una manera sintácticamente correcta (incluso si $P_w(w)$ no lo está —en cuyo caso el enunciado sería *verdadero*, puesto que no puede haber demostración de una expresión sintácticamente incorrecta). De hecho, debido a las traducciones a la aritmética que se supone hemos llevado a cabo, el enunciado de arriba es en realidad un enunciado *aritmético* relativo al número natural w (siendo la parte contenida entre paréntesis cuadrados un enunciado aritmético bien definido relativo a *dos* números naturales x y w). No resulta obvio que los enunciados puedan ser realmente codificados en la aritmética, aunque pueden serlo. Mostrar que tales enunciados pueden ser realmente codificados de esta forma es el principal "trabajo duro" incluido en la parte complicada del argumento de Gödel. Al igual que antes, *cuál* sea exactamente el enunciado aritmético dependerá de los detalles de numerosos sistemas, y dependerá en gran medida de la estructura detallada de los axiomas y reglas de nuestro sistema formal. Puesto que todo eso pertenece a la parte complicada, no nos interesaremos ahora por los detalles.

Hemos numerado todas las funciones proposicionales que dependen de una sola variable, de modo que la que acabamos de escribir debe tener asignado un número. Escribamos este número como k. Nuestra función proposicional es la k-ésima de la lista. Por consiguiente

$$\sim\exists x\ [\Pi_x \text{ demuestra } P_w(w)] = P_k(w)$$

Examinemos ahora esta función para el valor w particular: $w = k$. Tenemos

$$\sim\exists x\ [\Pi_x \text{ demuestra } P_k(k)] = P_k(k).$$

La proposición especifica $P_k(k)$ es un enunciado aritmético perfectamente bien definido (sintácticamente correcto). ¿Tiene una demostración dentro de nuestro sistema formal? ¿Tiene demostración su negación $\sim P_k(k)$? La respuesta a ambas preguntas debe ser "no". Podemos verlo examinando el *significado* subyacente en el procedimiento de Gödel. Aunque $P_k(k)$ es sólo una proposición aritmética, la hemos construido de modo que afirma lo que se ha escrito en el lado izquierdo: "no existe demostración, dentro del sistema, de la proposición $P_k(k)$". Si hemos sido cuidadosos al establecer nuestros axiomas y reglas de inferencia, y suponiendo que hayamos hecho bien nuestra numeración, entonces no puede haber ninguna demostración de esta $P_k(k)$ dentro del sistema. En efecto, si hubiera tal demostración, el significado del enunciado que $P_k(k)$ realmente afirma, a saber, que *no* existe demostración, sería falso, de modo que $P_k(k)$ tendría que ser falsa como proposición aritmética. Nuestro sistema formal no debería estar tan mal construido como para permitir que se demuestren proposiciones falsas. Por consiguiente, debe ser el caso que, de hecho, *no* hay demostración de $P_k(k)$. Pero esto es precisamente lo que $P_k(k)$ está tratando de decirnos. Por lo tanto, lo que afirma $P_k(k)$ debe ser un enunciado *verdadero*, de modo que $P_k(k)$ debe ser verdadera como proposición aritmética. Entonces hemos encontrado una proposición *verdadera* que *no tiene demostración dentro del sistema.*

¿Qué sucede con su negación $\sim P_k(k)$? Se concluye que haríamos mejor en no encontrar tampoco una demostración de esta otra. Acabamos de establecer que $\sim P_k(k)$ debe ser falsa (puesto que $P_k(k)$ es verdadera), pero se suponía que no podíamos demostrar proposiciones falsas dentro del sistema. Así, ni $P_k(k)$ ni $\sim P_k(k)$ son demostrables dentro de nuestro sistema formal. Esto establece el teorema de Gödel.

LA INTUICIÓN MATEMÁTICA

Nótese que aquí ha sucedido algo muy notable. Con frecuencia la gente considera el teorema de Gödel como algo negativo, algo que muestra las necesarias limitaciones del razonamiento matemático formalizado. Por muy abiertos que creamos ser, siempre habrá algunas proposiciones que escapan de la red. Pero ¿debería preocuparnos la proposición particular $P_k(k)$? En el curso del argumento anterior hemos establecido realmente que $P_k(k)$ es un enunciado *verdadero*. De algún modo nos las hemos arreglado para *ver* que $P_k(k)$ es verdadero pese al hecho de que esto no es formalmente demostrable dentro del sistema. Los formalistas matemáticos *estrictos* deberían estar realmente preocupados, ya que mediante su propio razonamiento hemos establecido que la noción formalista de "verdad" debe ser necesariamente incompleta. *Cualquiera* que sea el sistema formal (consistente) que se utilice para la aritmética, existen enunciados que consideramos verdaderos pero a los que no hemos asignado el valor de verdad **verdadero** mediante el procedimiento propuesto por los formalistas tal como se describió antes. Un formalista estricto quizá trataría de evitar esto no hablando para nada del concepto de verdad sino refiriéndose simplemente a la *demostrabilidad* dentro de algún sistema formal dado. Sin embargo, esto parece muy limitado. Utilizando este punto de vista, ni siquiera se puede estructurar el argumento de Gödel en la forma dada más arriba, puesto que partes esenciales hacen uso de razonamientos sobre lo que es verdadero y lo que no es verdadero.[2] Algunos formalistas adoptan un punto de vista mas "pragmático" al afirmar que no les preocupan enunciados tales como $P_k(k)$ debido a que son extremadamente complejos y sin interés como proposiciones de la aritmética. Estas personas alegarán:

> Sí, existe el extraño enunciado, tal como $P_k(k)$, para el que mi noción de demostrabilidad o **verdad** no coincide con su instintiva noción de verdad, pero estos enunciados no intervienen nunca en las matemáticas serias (al menos no en las que yo estoy interesado), ya que tales enunciados son absurdamente complicados y no naturales en las matemáticas.

Ciertamente se da el caso de que proposiciones como $P_k(k)$ sean (si se las escribe completas) extremadamente incómodas y extrañas en tanto que enunciados matemáticos acerca de números. Sin embargo, en años recientes se han expuesto algunos enunciados razonablemente sencillos de un carácter matemático muy aceptable, que son realmente equivalentes a proposiciones del tipo Gödel.[3] Éstas son indemostrables a partir de los axiomas normales de la aritmética, pero no obstante se deducen de una propiedad "evidentemente verdadera" que tiene el propio sistema axiomático.

La falta de interés que profesan los formalistas por la "verdad matemática" me parece un muy extraño punto de vista dentro de la filosofía de las matemáticas. Más aún: tampoco es realmente

---

[2] De hecho, el razonamiento en el teorema de Gödel puede presentarse de tal forma que no dependa de un concepto totalmente externo de *verdad* para proposiciones tales como $P_k(k)$. Sin embargo, depende todavía de una interpretación del "significado" real de *algunos* de los símbolos: en particular de que "~∃" *significa* realmente que "no existe ningún (número natural)... tal que...".

[3] En lo que sigue, las letras minúsculas representan números naturales, y las mayúsculas, conjuntos finitos de números naturales. Sea $m \rightarrow [n, k, r]$ la representación del enunciado: "si X es cualquier conjunto de m elementos de números naturales cuyos subconjuntos de k elementos son asignados a r cajas, entonces existe un 'gran' subconjunto Y de X de n elementos, tal que todos los subconjuntos de k elementos de Y entran en la misma caja" Cuando decimos "gran" subconjunto, significa que Y tiene más elementos que el número natural que es el elemento más pequeño de Y. Consideremos la proposición: "para cualquier elección de k, r y n existe un m_0 tal que, para todo m mayor que m_0, el enunciado $m \rightarrow [n, k, r]$ es siempre verdadero". J. Paris y L. Harrington (1977) han demostrado que esta proposición es equivalente a una proposición tipo Gödel para los axiomas estándar (Peano) de la aritmética, indemostrable a partir de éstos, aunque afirma algo acerca de estos axiomas que es obviamente verdadero (a saber, en este caso, que las proposiciones deducibles de los axiomas son ellas mismas verdaderas).

pragmático. Cuando los matemáticos llevan a cabo sus razonamientos, no desean tener que estar comprobando continuamente si sus argumentos pueden ser formulados o no en términos de los axiomas y reglas de inferencia de algún sistema formal complicado. Solamente necesitan estar seguros de que sus argumentos son vías válidas para llegar a la verdad. El argumento de Gödel es otro procedimiento válido, así que me parece que $P_k(k)$ es una verdad matemática tan buena como cualquiera que pueda obtenerse de forma más convencional utilizando los axiomas y las reglas de inferencia que puedan establecerse de antemano.

Un procedimiento que Gödel mismo sugiere es el siguiente: aceptemos que $P_k(k)$, que por el momento designaré simplemente por G_0, es una proposición perfectamente válida; por consiguiente, podemos añadirla a nuestro sistema como un axioma adicional. Por supuesto, nuestro nuevo sistema corregido tendrá su *propia* proposición de Gödel, llamémosla G_1, que de nuevo se ve que es un enunciado perfectamente válido acerca de números. En consecuencia, añadimos también G_1 a nuestro sistema. Esto nos da un nuevo sistema corregido que tendrá su propia proposición de Gödel G_2 (de nuevo perfectamente válida), que podemos añadir a continuación, obteniendo la siguiente proposición de Gödel G_3, que también añadimos, y así sucesivamente, repitiendo indefinidamente el proceso. ¿Qué sucede con el sistema resultante cuando nos permitimos utilizar la lista *completa* G_0, G_1, G_2, G_3,... como axiomas adicionales?

¿Pudiera ser que *ésta* fuera completa? Puesto que ahora tenemos un sistema de axiomas ilimitado (infinito) quizá no es evidente que sea aplicable el procedimiento de Gödel. Sin embargo, este continuo añadido de proposiciones de Gödel es un esquema perfectamente sistemático y puede ser reexpresado como un sistema lógico ordinario finito de axiomas y reglas de inferencia. Este sistema tendrá su propia proposición de Gödel, llamémosla G_ω, que puede ser añadida de nuevo y formar entonces la proposición de Gödel $G_{\omega+1}$ del sistema resultante. Repitiendo, como antes, obtenemos una lista G_ω, $G_{\omega+1}$, $G_{\omega+2}$, $G_{\omega+3}$,... de proposiciones, todas ellas enunciados perfectamente válidos acerca de números naturales, y que pueden ser todas añadidas a nuestro sistema formal. Esto es de nuevo perfectamente sistemático y nos lleva a un nuevo sistema que engloba a todo el lote; pero éste tendrá de nuevo su proposición de Gödel, llamémosla $G_{\omega+\omega}$ que podemos reescribir como $G_{\omega 2}$, todo el procedimiento puede empezar otra vez de modo que tendremos una nueva lista de axiomas $G_{\omega 2}$, $G_{\omega 2+1}$, $G_{\omega 2+2}$, etc., infinita pero sistemática, que conduce aún a un nuevo sistema y una nueva proposición de Gödel $G_{\omega 3}$. Repitiendo todo el procedimiento obtenemos $G_{\omega 4}$, y luego $G_{\omega 5}$ y así sucesivamente. Ahora bien, *este* procedimiento es completamente sistemático y tiene su propia proposición de Gödel $G_{\omega 2}$.

¿Termina esto alguna vez? En cierto sentido, no; pero nos conduce a algunas consideraciones matemáticas difíciles en las que no podemos entrar en detalle aquí. El procedimiento anterior fue discutido por Alan Turing en un artículo de 1939.[4] De hecho, y de forma muy notable, *cualquier* proposición verdadera (pero sólo cuantificada universalmente) de la aritmética puede ser obtenida mediante un proceso repetido de "gödelización" de este tipo. (Véase Feferman, 1988). Sin embargo, esto supone en cierto grado una petición de principio sobre cómo *decidimos*

[4] El título era "Sistemas de lógica basados en ordinales", y algunos lectores estarán familiarizados con la notación para los *números ordinales* de Cantor que he estado utilizando en los subíndices. La jerarquía de sistemas lógicos que se obtiene mediante el procedimiento que he descrito arriba está caracterizada por *números ordinales computables*.
Existen algunos teoremas matemáticos que son bastante naturales y fáciles de enunciar para los que, si intentamos demostrarlos utilizando las reglas estándar (Peano) de la aritmética, tendríamos que utilizar el procedimiento de "gödelización" anterior hasta un grado escandalosamente grande (extendiendo el procedimiento enormemente más allá de lo que he esbozado arriba). Las demostraciones matemáticas de estos teoremas no dependen en absoluto, como otras, de algún vago o cuestionable razonamiento que pareciera estar fuera de los procedimientos de la argumentación matemática normal. Véase Smorynski (1983).

realmente si una proposición es verdadera o falsa. El punto crítico, en cada paso, consiste en establecer cómo codificar el añadido de una familia infinita de proposiciones de Gödel que proporciona un simple axioma adicional (o un número finito de axiomas). Esto requiere que nuestra familia infinita pueda ser sistematizada de alguna forma algorítmica. Para estar seguros de que tal sistematización hace *correctamente* lo que se supone que hace, necesitamos utilizar *intuiciones* procedentes del exterior del sistema —igual que lo hicimos en primer lugar para ver que $P_k(k)$ era una proposición verdadera. Son estas intuiciones las que no pueden ser sistematizadas— y, en realidad, deben estar fuera de *cualquier* acción algorítmica.

La intuición a partir de la que concluimos que la proposición de Gödel $P_k(k)$ es realmente un enunciado verdadero de la aritmética es un ejemplo de un tipo general de procedimiento que los lógicos conocen como *principio de reflexión*: así, "reflejando" el *significado* del sistema de axiomas y reglas de inferencia, y convenciéndonos de que éstos proporcionan realmente vías válidas para alcanzar las verdades matemáticas, podemos ser capaces de codificar esta intuición en nuevos enunciados matemáticos verdaderos que no eran deducibles de aquellos mismos axiomas y reglas. La inferencia de la verdad de $P_k(k)$ como se resumió arriba, dependía de un principio semejante. Otro principio de reflexión, importante para el argumento original de Gödel (que no se ha mencionado), descansa en la deducción de nuevas verdades matemáticas a partir del hecho de que un sistema axiomático, que creemos es válido para obtener verdades matemáticas, es realmente *consistente*. Los principios de reflexión implican con frecuencia razonamientos sobre conjuntos infinitos, y hay que ser cuidadosos al utilizarlos de modo que no se esté demasiado cerca del tipo de argumento que pudiera conducirnos a una paradoja tipo Russell. Los principios de reflexión proporcionan la propia antítesis del razonamiento formalista. Si se es cuidadoso, nos permiten salir fuera de los rígidos confinamientos de cualquier sistema formal y obtener nuevas intuiciones matemáticas que no parecían disponibles antes. Podría haber muchos resultados perfectamente aceptables en nuestra literatura matemática cuyas demostraciones requieran intuiciones que quedan lejos de los axiomas y reglas originales de los sistemas formales estándar de la aritmética. Todo esto muestra que los procedimientos mentales mediante los que los matemáticos llegan a sus juicios de verdad no están simplemente enraizados en los procedimientos de algún sistema formal específico. *Vemos* la validez de la proposición de Gödel $P_k(k)$ aunque no podamos derivarla de los axiomas. El tipo de "visión" que está implicada en un principio de reflexión requiere un acto de intuición esencial matemática que no es el resultado de las operaciones puramente algorítmicas que pudieran ser codificadas en algún sistema matemático formal. Volveremos a tratar este asunto en el capítulo X.

El lector puede notar una cierta similitud entre el argumento que establece la verdad, aunque "indemostrabilidad", de $P_k(k)$ y el argumento de la paradoja de Russell. También hay una similitud con el argumento de Turing que establece la no existencia de una máquina de Turing que resuelva el problema de la detención. Estas similitudes no son accidentales. Existe un fuerte hilo conductor histórico entre los tres. Turing encontró su argumento después de estudiar el trabajo de Gödel. El mismo Gödel conocía bien la paradoja de Russell y fue *capaz* de transformar el razonamiento paradójico de este tipo, que lleva demasiado lejos el uso de la lógica, en un argumento matemático válido. (Todos estos argumentos tienen su origen en el "corte diagonal" de Cantor, descrito en el capítulo anterior.)

¿Por qué deberíamos aceptar los argumentos de Gödel y Turing cuando hemos tenido que rechazar el razonamiento que conduce a la paradoja de Russell? Los primeros son mucho más nítidos y son irreprochables como argumentos matemáticos, mientras que la paradoja de Russell

descansa en razonamientos más vagos que involucran conjuntos "inmensos". Pero debe admitirse que las diferencias no son realmente tan claras como a uno le gustaría que fuesen. El intento de clarificar estas diferencias era un motivo poderoso tras la idea global del formalismo. El argumento de Gödel muestra que el punto de vista del formalismo estricto es insostenible, pero no nos lleva a un punto de vista alternativo completamente fiable. Para mí, el resultado final sigue sin estar resuelto. El método que se adopta* de hecho en las matemáticas contemporáneas para evitar el tipo de razonamiento con conjuntos "inmensos" que conduce a la paradoja de Russell no es enteramente satisfactorio. Además, aún suele establecerse en términos típicamente formalistas o, alternativamente, en términos que no nos dan una confianza plena de que no puedan aparecer contradicciones.

De cualquier modo, me parece que una consecuencia evidente del argumento de Gödel es que el concepto de verdad matemática no puede ser encapsulado en ningún esquema formalista. La verdad matemática es algo que trasciende el mero formalismo. Esto es quizá evidente aún sin el teorema de Gödel. En efecto, ¿cómo vamos a decidir qué axiomas o reglas de inferencia adoptar en un caso cualquiera cuando tratamos de establecer un sistema formal? Nuestra guía para la decisión de las reglas que vamos a adoptar debe ser siempre nuestra comprensión intuitiva de lo que es "evidentemente verdadero", dados los "significados" de los símbolos del sistema. ¿Cómo vamos a decidir qué sistemas formales son razonables para ser adoptados —es decir, que están de acuerdo con nuestras ideas intuitivas sobre "evidencia" y "significado"— y cuáles no? Ciertamente, la noción de consistencia no es adecuada para ello. Podemos tener muchos sistemas consistentes que no son "razonables" en este sentido, en los que los axiomas y reglas de inferencia tienen significados que rechazaríamos como falsos, o quizá no tengan significado en absoluto. Por lo tanto, "evidencia" y "significado" son conceptos que seguirían siendo necesarios aún sin el teorema de Gödel.

Sin embargo, aún sin el teorema de Gödel hubiera sido posible imaginar que las nociones intuitivas de "evidencia" y "significado" podrían haber sido empleadas sólo una vez y para siempre, simplemente para establecer el sistema formal en primer lugar, y prescindir de ellas en lo sucesivo como parte de un argumento matemático claro para determinar la verdad. Entonces, según el punto de vista formalista, estas "vagas" nociones intuitivas hubieran tenido algún papel que jugar como parte del pensamiento matemático *preliminar*, como una guía hacia el descubrimiento del argumento formal apropiado, pero no desempeñarían ningún papel en la demostración real de la verdad matemática. El teorema de Gödel muestra que este punto de vista no es sostenible en una filosofía de los fundamentos de las matemáticas. La noción de verdad matemática va más allá del concepto global de formalismo. Hay algo absoluto e "infuso" en la verdad matemática. De esto trata el platonismo matemático, como se discutió al final del anterior capítulo. Cualquier sistema formal concreto tiene una cualidad provisional y "de factura humana". Tales sistemas pueden desempeñar papeles muy valiosos en las discusiones matemáticas, pero sólo pueden proporcionar una guía parcial (o aproximada) a la verdad. La verdad matemática real va más allá de las simples construcciones humanas.

* Se hace una distinción entre "conjuntos" y "clases", en la que se permite que los conjuntos se agrupen para formar otros conjuntos o tal vez clases, pero *no* se permite que las clases se agrupen para formar colecciones mayores de cualquier tipo, siendo consideradas como "demasiado grandes" para esto. Sin embargo, no existe ninguna regla para decidir cuándo se permite que una colección pueda ser considerada como un conjunto o cuándo deba considerarse necesariamente que sólo es una clase, aparte de la regla circular que establece que los conjuntos son aquellas colecciones que pueden agruparse para formar otras colecciones.

¿PLATONISMO O INTUICIONISMO?

He señalado dos escuelas contrarias de filosofía matemática, que se inclinan fuertemente hacia al platonismo antes que hacia el punto de vista formalista. En realidad he sido bastante simplista en mis distinciones, porque este punto presenta muchos matices. Por ejemplo, partiendo del criterio de platonismo, uno puede cuestionarse si los objetos del pensamiento matemático poseen algún tipo de "existencia" real o si es sólo el concepto de "verdad" matemática el que es absoluto. No he querido plantear aquí estas distinciones. A mi modo de ver, el carácter absoluto de la verdad matemática y la existencia platónica de los conceptos matemáticos son esencialmente la misma cosa. La "existencia" que debe atribuirse al conjunto de Mandelbrot, por ejemplo, es una característica de su naturaleza "absoluta". El que un punto del plano de Argand pertenezca o no al conjunto de Mandelbrot es una cuestión absoluta, independiente de qué matemático, o qué computadora, lo esté examinando. Es la "independencia del sujeto" del conjunto de Mandelbrot la que le confiere su existencia platónica. Además, sus detalles más finos quedan fuera del alcance de las computadoras. Estos dispositivos sólo pueden darnos aproximaciones a una estructura que tiene en sí misma una existencia más profunda e "independiente de la computadora". Reconozco, sin embargo, que puede haber muchos otros puntos de vista a propósito de esta cuestión, pero no tenemos aquí que preocuparnos por estas diferencias.

Hay también diferencias de punto de vista sobre hasta qué extremo estamos dispuestos a llevar nuestro platonismo —si realmente uno afirma ser un platónico—. El propio Gödel era un gran platónico. Los tipos de enunciados matemáticos que he estado considerando hasta ahora son más bien "tibios" tal como van las cosas.[5] Pueden surgir enunciados mucho más controvertidos, particularmente en la teoría de conjuntos. Cuando se consideran todas las ramificaciones de la teoría de conjuntos se tropieza con conjuntos tan desmesuradamente enormes, y construidos de manera tan vaga, que incluso un decidido platónico como yo puede honestamente empezar a dudar que su existencia, o inexistencia, sea realmente algo "absoluto".[6] Puede llegar un momento en que los conjuntos tengan una definición tan intrincada y conceptualmente dudosa que la cuestión de la verdad o falsedad de enunciados matemáticos relativos a ellos pueda empezar a adquirir algo de la cualidad de "cuestión de opinión" en lugar de la de "infusa". El que uno esté preparado para llevar el platonismo hasta sus últimas consecuencias, junto con Gödel, y exigir que la verdad o falsedad de los enunciados matemáticos relativos a tan enormes conjuntos sea siempre algo absoluto o "platónico", o bien se detenga en algún punto anterior y exija una verdad o falsedad absoluta sólo cuando los conjuntos son razonablemente constructivos y no tan desmesuradamente enormes, no es un asunto que tenga aquí gran relevancia para nuestra discusión. Los conjuntos (finitos o infinitos) que tendrán importancia para nosotros son ridículamente minúsculos comparados con aquellos a los que me acabo de referir. Por ello las diferencias entre estas diversas visiones platónicas no nos afectan grandemente.

---

[5] La hipótesis del continuo que fue mencionada en el capítulo III, (y que establece que $C = \aleph_1$) es el enunciado matemático más "extremo" que hemos encontrado aquí (aunque con frecuencia se consideran enunciados mucho más extremos que éste). La hipótesis del continuo tiene un interés adicional debido a que el propio Gödel, junto con Paul J. Cohen, estableció que esta hipótesis es en realidad *independiente* de los axiomas y reglas de inferencia estándar de la teoría de conjuntos. Por consiguiente, la actitud de cada uno hacia el status de la teoría del continuo distingue entre los puntos de vista formalista y platónico. Para un formalista la hipótesis del continuo es indecidible puesto que no puede ser establecida ni refutada utilizando el sistema formal estándar (Zermelo-Frankel), y no tiene sentido llamarla "verdadera" o "falsa". Sin embargo, para un buen platónico la hipótesis del continuo es realmente o verdadera o falsa, aunque establecer cuál de los dos es el caso requeriría alguna forma nueva de razonamiento, que va incluso más allá del empleo de Proposiciones tipo Gödel para el sistema formal de Zermelo-Frankel. (El propio Cohen [1966] sugirió un principio de reflexión que muestra a la hipótesis del continuo como obviamente falsa".)

[6] Para un informe vivo y no técnico sobre estas cuestiones, véase Rucker (1984).

Existen, no obstante, otros puntos de vista matemáticos, como el que se conoce como *intuicionismo* (y que otros llaman *finitismo*), que van al otro extremo en donde se rechaza la existencia consumada de cualquier conjunto infinito.* El intuicionismo fue iniciado en 1924 por el matemático holandés L. E. J. Brouwer como una respuesta alternativa —diferente del formalismo— a las paradojas (como la de Russell) que pueden aparecer cuando se utilizan demasiado libremente los conjuntos infinitos en el razonamiento matemático. Las raíces de este punto de vista pueden rastrearse hasta Aristóteles, que había sido discípulo de Platón pero había rechazado sus puntos de vista acerca de la existencia absoluta de las entidades matemáticas y sobre la aceptabilidad de conjuntos infinitos. Según el intuicionismo, los conjuntos (infinitos o no) no deben pensarse como si tuvieran "existencia" por sí mismos, sino que deben pensarse simplemente en términos de las reglas por las que se puede determinar su pertenencia a ellos.

Una característica distintiva del intuicionismo de Brouwer es el rechazo de la "ley del tercio excluido". Esta ley afirma que la negación de la negación de un enunciado es equivalente a la afirmación de dicho enunciado. (En símbolos: $\sim(\sim P) \Leftrightarrow P$, una relación que encontramos antes.)**. Aristóteles no se hubiera sentido feliz con la negación de algo tan lógicamente "obvio" como esto. En términos ordinarios de "sentido común", la ley del tercio excluido puede considerarse como una verdad evidente: si es falso que algo no es verdadero, entonces este algo es ciertamente verdadero. (Esta ley es la base del método matemático de *reductio ad absurdum*). Pero los intuicionistas se creen capaces de negar esta ley. Esto se debe básicamente a que adoptan una actitud diferente hacia el concepto de *existencia*, exigiendo que se presente una construcción (mental) definida antes de aceptar que un objeto matemático existe realmente. Por ello, para un intuicionista "existencia" significa "existencia constructiva". En un argumento matemático que procede por *reductio ad absurdum* uno desarrolla alguna hipótesis con la intención de mostrar que sus consecuencias conducen a una contradicción, contradicción que proporciona la deseada demostración de que la hipótesis en cuestión es falsa. La hipótesis podría tomar la forma de un enunciado acerca de que una entidad matemática con ciertas propiedades requeridas no existe. Cuando esto conduce a una contradicción, uno infiere, en *matemáticas ordinarias*, que la entidad requerida existe realmente. Pero tal argumento, por sí mismo, no proporciona medios para *construir* efectivamente tal entidad. Para un intuicionista, este tipo de existencia no es existencia en absoluto; y es en este sentido en el que rechazan aceptar la ley del tercio excluido y el método de *reductio ad absurdum*. En realidad, Brouwer estaba profundamente insatisfecho con tal "existencia" no constructiva.[7] Sin una construcción real,

---

* El intuicionismo fue llamado así debido a que se suponía que reflejaba el pensamiento humano.

** La ley $\sim(\sim P) \Leftrightarrow P$ se conoce más frecuentemente como "ley de la doble negación", reservándose el nombre de "ley del tercio excluido" para la ley $P \vee \sim P$ (o P es verdadera o P es falsa). Ambas son equivalentes en la lógica clásica ordinaria, aunque no se puede decir lo mismo en la lógica intuicionista que no es veritativo-funcional ni admite la interdefinición de las conectivas. Obviamente, la lógica intuicionista niega ambas leyes. [Nota del traductor.]

[7] El propio Brouwer parece haber partido de esta línea de razonamiento debido en parte a las quejas y críticas acerca de una "no constructividad" en su demostración de uno de sus propios teoremas: el teorema del punto fijo de Brouwer de la topología. El teorema afirma que si se toma un disco —es decir, un círculo junto con su interior— y se deforma de una manera continua hacia el interior de la región en la que estaba situado inicialmente, entonces existe al menos un punto del disco —llamado punto fijo— que termina exactamente donde empezó. Podemos no tener idea de cuál es exactamente este punto, o de si pudiera haber varios de estos puntos; lo que afirma el teorema es simplemente la *existencia* de alguno de estos puntos. (Como son los teoremas de existencia en matemáticas, éste es en verdad completamente "constructivo". De un orden diferente de no constructividad son los teoremas de existencia que dependen de lo que se conoce como el "axioma de elección" o "lema de Zorn" (cfr. Cohén, 1966: Rucker, 1984.) En el caso de Brouwer la dificultad es similar a lo siguiente: si f es una función continua real de variable real, que toma valores positivos y negativos, encontrar el punto en el que f se anula. El procedimiento usual supone la bisección repetida del intervalo en el que f cambia de signo, pero puede no ser "constructivo", en el sentido requerido por Brouwer, el decidir si los valores intermedios de f son positivos, negativos o cero.

aseguraba, dicho concepto de existencia no tiene significado. En la lógica brouweriana, de la falsedad de la no existencia de un objeto no se puede deducir que exista realmente. En mi opinión, aunque es encomiable buscar la constructividad en la existencia matemática, el punto de vista del intuicionismo de Brouwer es demasiado extremo. Brouwer expuso sus ideas por primera vez en 1924, más de diez años antes del trabajo de Church y Turing. Ahora que el concepto de constructividad —en términos de la idea de computabilidad de Turing— puede estudiarse dentro del marco *convencional* de la filosofía matemática, no hay necesidad de llegar a los extremos a los que Brouwer quería llevarnos. Podemos discutir la constructividad como un tema separado de la cuestión de la existencia matemática. Si seguimos con el intuicionismo, debemos negarnos el uso de tipos de argumentos muy poderosos dentro de las matemáticas, y el tema se vuelve de algún modo agobiante e impotente.

No deseo extenderme sobre las diversas dificultades o absurdos aparentes a que nos lleva el punto de vista intuicionista, aunque quizá sea útil que mencione sólo algunos de los problemas. Un ejemplo citado a menudo por Brouwer se refiere a la expansión decimal de π:

$$3.141592653589793....$$

¿Existe una serie de veinte sietes consecutivos en algún lugar de esta expansión, es decir

$$\pi = 3.141592653589793...\ 77777777777777777777...,$$

o no existe tal serie? En términos matemáticos ordinarios, todo lo que podemos decir por ahora es que o existe o no existe, y no sabemos cuál de las dos posibilidades ocurre. Este parecería ser un enunciado bastante inofensivo. Sin embargo, los intuicionistas negarán que sea válido decir "o existe una serie de veinte sietes consecutivos en algún lugar de la expansión decimal de π o no existe" —a menos que se haya establecido (de alguna manera constructiva aceptable para los intuicionistas) que realmente existe tal serie, o bien se haya establecido que no existe ninguna. Un cálculo directo bastaría para mostrar que una serie de veinte sietes consecutivos existe en algún lugar de la expansión decimal de π, pero se necesitaría alguna especie de teorema matemático para establecer que no hay tal serie. Ninguna computadora ha llegado aún lo suficientemente lejos en el cálculo de π *para* determinar que existe una serie semejante. Lo que uno esperaría sobre bases probabilísticas es que tal serie exista, pero aún si una computadora pudiera calcular consistentemente dígitos a un ritmo de, digamos, 10^{10} por segundo ¡llevaría probablemente un tiempo del orden de entre cien y mil años encontrar la serie! Me parece mucho más probable que, antes que por computación directa, la existencia de dicha serie sea establecida matemáticamente (probablemente como un corolario de algún resultado mucho más poderoso e interesante), aunque quizá no de una forma aceptable para los intuicionistas.

Este problema concreto no tiene interés matemático real. Sólo se da como un ejemplo que es fácil de explicar. En la forma de intuicionismo extremo de Brouwer, él afirmaría que, en el momento presente, la afirmación "existe una serie de veinte sietes consecutivos en algún lugar de la expansión decimal de π" no es verdadera ni falsa. Si, en alguna fecha posterior, se estableciera el resultado correcto en uno u otro sentido, mediante computación o mediante demostración matemática (intuicionista), entonces la afirmación se convertiría en "verdadera" o "falsa", según sea el caso. Un ejemplo similar sería el "último teorema de Fermat". Según el intuicionismo

Una versión intuitiva de este teorema es que si se quita un mantel que recubre una mesa circular, se arruga sin desgarrarlo y se tira así sobre la mesa, habrá al menos un punto del mantel que no habrá cambiado de lugar (citado del "Diccionario de Matemáticas", dirigido por Francois le Lionnais, Editorial Akal). [N. del T.]

extremo de Brouwer, éste tampoco es verdadero ni falso, pero podría llegar a ser uno u otro en alguna fecha posterior. Para mí, tal subjetividad y dependencia temporal de la verdad matemática es inadmisible. Es, en verdad, una cuestión muy subjetiva la de si, o cuando, un resultado matemático puede aceptarse como oficialmente "demostrado". La verdad matemática no debe descansar en semejantes criterios sociodependientes. Además, tener un concepto de la verdad matemática que cambia con el tiempo, es, al menos, muy incómodo e insatisfactorio para unas matemáticas que esperamos puedan ser empleadas fiablemente en una descripción del mundo físico. No todos los intuicionistas adoptarían una posición tan estricta como la de Brouwer. De todas formas, el punto de vista intuicionista es decididamente incómodo, incluso para los que simpatizan con los objetivos del constructivismo. Pocos matemáticos actuales se alinean incondicionalmente con el intuicionismo, aunque sólo sea porque es muy restrictivo respecto a los tipos de razonamiento matemático que nos permite utilizar.

He descrito brevemente las tres corrientes principales de la filosofía matemática actual: formalismo, platonismo e intuicionismo. No he ocultado mis fuertes simpatías por el punto de vista platónico de que la verdad matemática es absoluta, externa y eterna, y no se basa en criterios hechos por el hombre; y que los objetos matemáticos tienen una existencia intemporal por sí mismos, independiente de la sociedad humana o de objetos físicos particulares. He tratado de presentar mis argumentos a favor de este punto de vista en esta sección, en la sección anterior y al final del capítulo III. Espero que el lector esté preparado para seguir conmigo en este camino. Será importante para mucho de lo que encontraremos más adelante.

TEOREMAS TIPO GÖDEL A PARTIR DEL RESULTADO DE TURING

En mi presentación del teorema de Gödel he omitido muchos detalles, y también he dejado de lado lo que históricamente fue la parte quizá más importante de su argumento: la que se refiere a la "indecidibilidad" de la consistencia de los axiomas. Mi propósito aquí *no* ha sido el de subrayar este "problema de la demostrabilidad de la consistencia de los axiomas", tan importante para Hilbert y sus contemporáneos, sino mostrar que, utilizando nuestras intuiciones de los significados de las operaciones en cuestión, se *ve* claramente que una proposición de Gödel específica —ni demostrable ni indemostrable utilizando los axiomas y reglas del sistema formal considerado— es una proposición *verdadera*.

He mencionado que Turing desarrolló su propio argumento posterior que establecía la insolubilidad del problema de la detención después de haber estudiado el trabajo de Gödel. Los dos argumentos tienen mucho en común y, de hecho, los aspectos claves del resultado de Gödel pueden derivarse directamente utilizando el método de Turing. Veamos cómo funciona esto, y de ahí obtendremos una opinión algo diferente sobre lo que hay detrás del teorema de Gödel.

Una propiedad esencial de un sistema matemático formal es que debería ser una cuestión computable el decidir si una cadena dada de símbolos constituye o no una demostración, dentro del sistema, de un enunciado matemático dado. Después de todo, la idea general en la formalización de la noción de demostración matemática es que no habrá que hacer juicios posteriores sobre lo que es un razonamiento válido y lo que no lo es. Debe ser posible verificar de una forma completamente mecánica y previamente determinada si una supuesta demostración es o no una demostración; es decir, debe haber un *algoritmo* para verificar demostraciones. Por

otra parte, no exigimos que deba ser necesariamente una cuestión algorítmica el *encontrar* demostraciones (o refutaciones) de enunciados matemáticos propuestos.

Resulta, de hecho, que siempre *existe* un algoritmo para encontrar una demostración dentro de cualquier sistema matemático formal en el que dicha demostración exista. En efecto, supongamos que nuestro sistema está formulado en términos de algún lenguaje simbólico, que a su vez es expresable en términos de algún "alfabeto" finito de símbolos. Como antes, ordenemos *lexicográficamente* nuestras cadenas de símbolos, lo que significa, recordemos, hacer una ordenación alfabética para cada longitud de cadena dada, ordenando primero alfabéticamente todas las cadenas de longitud uno, a continuación las de longitud dos, luego las de longitud tres y así sucesivamente . De esta forma tenemos todas las demostraciones correctamente construidas ordenadas numéricamente según un esquema lexicográfico. Teniendo nuestra lista de demostraciones, tenemos también una lista de todos los *teoremas* del sistema formal, pues los teoremas son precisamente las proposiciones que aparecen en la última línea de las demostraciones correctamente construidas. El listado es perfectamente computable; en efecto, podemos considerar la lista lexicográfica de *todas* las cadenas de símbolos del sistema, tengan o no sentido como demostraciones, y luego comprobar la primera cadena con nuestro algoritmo de verificar demostraciones para ver si es una demostración y descartarla si no lo es; luego comprobamos la segunda de la misma manera y la descartamos si no es una demostración, y luego la tercera, luego la cuarta y así sucesivamente. De este modo, si existe una demostración acabaremos por encontrarla en algún lugar de la lista.

Por consiguiente, si Hilbert hubiera tenido éxito en encontrar su sistema matemático —un sistema de axiomas y reglas de inferencia lo bastante fuerte para que podamos decidir, mediante demostración formal, la verdad o falsedad de cualquier proposición matemática correctamente formulada dentro del sistema— entonces *existiría* un método algorítmico general para decidir la verdad de cualquiera de estas proposiciones. ¿Por qué es esto así? Lo es porque si, mediante el procedimiento esbozado arriba, podemos llegar a encontrar la proposición que estamos buscando como la línea final en la demostración, entonces hemos *demostrado* esa proposición. Si, por el contrario, llegamos a encontrar la *negación* de nuestra proposición como la línea final, entonces la hemos *refutado*. Si el esquema de Hilbert fuera completo siempre ocurriría una u otra de estas eventualidades (y, si fuera consistente, nunca ocurrirían las dos juntas). Así, nuestro procedimiento mecánico terminará siempre en algún paso y tendríamos un algoritmo universal para decidir la verdad o falsedad de todas las proposiciones del sistema. Esto contradiría el resultado de Turing, como se presentó en el capítulo II, en el sentido de que no existe algoritmo general para decidir proposiciones matemáticas. Por consiguiente hemos demostrado, en efecto, el teorema de Gödel de que *ningún* esquema del tipo propuesto por Hilbert puede ser completo en el sentido que hemos estado discutiendo.

En realidad, el teorema de Gödel es más concreto que esto, puesto que el tipo de sistema formal en el que Gödel estaba interesado debía ser apropiado sólo para proposiciones de la aritmética, y no para proposiciones de las matemáticas en general. ¿Podemos hacer que todas las operaciones de las máquinas de Turing estén relacionadas sólo con la aritmética? ¿Podemos expresar todas las funciones *computables* de números naturales (es decir, las funciones recursivas o algorítmicas —los resultados de la acción de una máquina de Turing—) en términos de aritmética ordinaria? Podemos, pero no del todo. Necesitamos añadir una operación extra a las reglas estándar de la aritmética y la lógica (incluso $\exists$ y $\forall$). Esta operación selecciona

"el menor número natural x tal que $K(x)$ es verdadero",

donde $K(\,)$ es cualquier función proposicional dada calculable aritméticamente —para la que se supone que *existe* tal número, es decir, que $\exists x\ [K(x)]$ es verdadero. (Si no existiera tal número, nuestra operación seguiría "actuando indefinidamente"* para localizar el no existente x requerido.) En cualquier caso, el argumento precedente establece, sobre la base del resultado de Turing, que el programa de Hilbert (de reducir ramas enteras de las matemáticas a cálculos dentro de algún sistema formal) es insostenible.

Tal como están las cosas, este método no prueba que tengamos una proposición de Gödel (como $P(k)$ que es *verdadera*, aunque no demostrable dentro del sistema. Sin embargo, si recordamos el argumento dado en el capítulo II sobre "cómo superar a un algoritmo" , veremos que podemos hacer algo parecido. En dicho argumento pudimos probar que, dado un algoritmo para determinar si una máquina de Turing se detiene, podemos plantear una acción de una máquina de Turing que *nosotros* vemos que no se detiene, aunque el algoritmo no puede hacerlo. (Recuérdese que insistíamos en que el algoritmo debe informarnos correctamente cuándo se *detendrá* la acción de una máquina de Turing, si bien a veces puede fallar al no decirnos que tal máquina no se detendrá, porque seguirá funcionando indefinidamente.) De este modo, como sucedía en la situación anterior con el teorema de Gödel, tenemos una proposición que podemos *ver*, mediante una intuición, que debe ser *verdadera* (sin detener la acción de la máquina de Turing), aunque la acción algorítmica dada no sea capaz de decírnoslo.

CONJUNTOS RECURSIVAMENTE ENUMERABLES

Hay una manera gráfica de describir los ingredientes básicos de los resultados de Turing y Gödel: mediante la *teoría de conjuntos*, que nos permite alejarnos de las descripciones arbitrarias en términos de simbolismos específicos o sistemas formales, y pone de manifiesto los resultados esenciales. Consideraremos sólo conjuntos (finitos o infinitos) de *números naturales* 0, 1, 2, 3, 4,..., de manera que examinemos colecciones de éstos, tales como {4, 5, 8}, {0, 57, 100003}, {6}, {0}, {1, 2, 3, 4,..., 9999}, {1, 2, 3, 4,...}, {0, 2, 4, 6, 8,...}, o incluso el conjunto total $\mathbb{N}$ ={0, 1, 2, 3, 4,...} o el conjunto vacío ∅= { }. Nos interesaremos sólo en cuestiones de *computabilidad*, a saber: ¿qué tipos de conjuntos de números naturales pueden ser generados mediante algoritmos y cuáles no?

Podemos pensar, si así lo deseamos, que un número natural n denota una cadena específica de símbolos en un sistema formal particular. Esta sería la "n-ésima" cadena de símbolos, llamémosla Q_n, de acuerdo con una cierta ordenación lexicográfica de las proposiciones del sistema (expresadas de forma "sintácticamente correcta"). Entonces, cada número natural representa una proposición. El conjunto de *todas* las proposiciones del sistema formal estará representado por el conjunto total $\mathbb{N}$ y, por ejemplo, los *teoremas* del sistema formal serán imaginados como un conjunto más pequeño de números naturales, digamos el conjunto P. Pero debido a que los detalles de un sistema particular de numeración no son importantes, lo único que necesitaremos para establecer una correspondencia entre los números naturales y las proposiciones será conocer un algoritmo para cada proposición Q_n (escrita en la notación

* Es esencial aceptar que se permita que puedan darse estas desafortunadas posibilidades, de modo que podamos describir potencialmente *cualquier* operación algorítmica. Recuérdese que para describir máquinas de Turing en general, debemos admitir la existencia de máquinas de Turing que no se detienen nunca.

simbólica adecuada), a partir de su correspondiente número natural n, y conocer otro algoritmo para obtener n a partir de Q_n. Con estos dos algoritmos conocidos estaremos en libertad de *identificar* el conjunto de los números naturales $\mathbb{N}$ con el conjunto de las proposiciones de un sistema formal específico.

Escojamos un sistema formal que sea consistente y suficientemente amplio para incluir todas las acciones de las máquinas de Turing —y, más aún, "razonable" en el sentido de que sus axiomas y reglas de inferencia pueden asumirse como "autoevidentemente *verdaderos*".

Algunas de las proposiciones Q_0, Q_1, Q_2, Q_3,... del sistema formal tendrán *demostraciones* dentro del sistema. Esas proposiciones "demostrables" tendrán números que constituirán en $\mathbb{N}$ el conjunto P de "teoremas" de que hablamos antes. Ya hemos visto que existe un *algoritmo* para generar, una detrás de otra, todas las proposiciones con demostraciones en algún sistema formal dado, y, como se esbozó antes, la "n-ésima demostración" Π_n se obtiene algorítmicamente a partir de n: todo lo que tenemos que hacer es mirar la última línea de la n-ésima demostración para encontrar la "n-ésima proposición demostrable dentro del sistema", es decir, el n-ésimo "teorema". De este modo, tenemos un algoritmo para generar los elementos de P uno detrás de otro (quizá con repeticiones, pero ello no marca diferencia).

Un conjunto como P, que puede generarse mediante un algoritmo, se llama *recursivamente enumerable*. Nótese que el conjunto de las proposiciones que son indemostrables dentro del sistema —es decir, las proposiciones cuya negación es demostrable—, es también recursivamente enumerable, pues basta con enumerar las demostraciones demostrables y tomar sus negaciones a medida que avanzamos. Hay muchos otros subconjuntos de $\mathbb{N}$ que son recursivamente numerables, y no necesitados hacer referencia a nuestro sistema formal para definirlos. Ejemplos sencillos de conjuntos recursivamente enumerables son el conjunto de los números pares

$$\{0,2,4,6,8,...\},$$

el conjunto de los cuadrados

$$\{0,1,4,9,16,...\},$$

y el conjunto de los primos

$$\{2,3,5,7,11,...\}.$$

Evidentemente, podemos generar cada uno de estos conjuntos por medio de un algoritmo. En cada uno de los tres ejemplos sucederá *también* que el *complemento* del conjunto, es decir, el conjunto de los números naturales que *no* están en el conjunto, es recursivamente enumerable. Los conjuntos complementarios en estos tres casos son, respectivamente:

$$\{1,3,5,7,9,...\},$$

$$\{2,3,5,6,7,8,10,...\},$$

y

$$\{0,1,4,6,8,9,10,12...\}.$$

Sería un asunto sencillo proporcionar también un algoritmo para estos conjuntos complementarios y hasta para cualquier número natural n dado, porque es posible determinar

algorítmicamente si se trata o no de un número par, si es o no un cuadrado, o si es o no un número primo. Esto nos proporciona un algoritmo para generar *ambos*, el conjunto y el conjunto complementario, porque podemos recorrer los números naturales uno a uno y saber en cada caso si pertenece al conjunto original o al conjunto complementario. El conjunto que tiene la propiedad de que tanto él como su complementario son recursivamente enumerables, se denomina conjunto *recursivo, y* por eso el complemento de un conjunto recursivo es también un conjunto recursivo.

Ahora bien, ¿existen conjuntos que sean recursivamente numerables pero *no* recursivos? Hagamos una pausa por un momento, para señalar lo que esto supone. Dado que los elementos de tal conjunto pueden ser generados mediante un algoritmo, tendremos un medio para establecer si un elemento sospechoso de pertenecer al conjunto —y que, supongamos por el momento, sí pertenece al conjunto— *está*, en efecto, en el conjunto. Todo lo que tenemos que hacer es permitir que nuestro algoritmo actúe sobre los elementos del conjunto hasta que finalmente llene al elemento particular que estamos examinando.

Pero supongamos que nuestro elemento sospechoso *no* está realmente en el conjunto. En este caso nuestro algoritmo no nos dará resultado porque seguirá actuando indefinidamente sin llegar a una decisión. Para eso necesitaríamos un algoritmo que generara el conjunto *complementario.* Si *este* algoritmo descubriera a nuestro sospechoso, tendríamos la certeza de que el elemento no está en el conjunto. Con ambos algoritmos podríamos trabajar, alternándolos y atrapando a nuestro sospechoso. Sin embargo, tal situación feliz es la que ocurre con un conjunto *recursivo.* En contraprueba, consideremos ahora un conjunto que es sólo recursivamente numerable, *no* recursivo: nuestro algoritmo para generar el conjunto complementario no existe. Se nos presenta así una situación en la que, para un elemento *perteneciente* al conjunto, podemos establecer con algoritmos que efectivamente *está* en el conjunto, pero no podemos, con tal método, solucionar el problema (con garantías) para elementos que *no* están en el conjunto.

¿Se presenta alguna vez tan curiosa situación? ¿Existen conjuntos recursivamente numerables que no son recursivos? ¿Qué hay sobre el conjunto *P*? ¿Es *este* un conjunto recursivo? Sabemos que es recursivamente numerable, pero hasta aquí desconocemos si el conjunto complementario es también recursivamente numerable.

No lo es. ¿Cómo podemos decir esto? Hemos supuesto que las acciones de las máquinas de Turing figuran entre las operaciones permitidas dentro de nuestro sistema formal. Denotemos por T_n la n-ésima máquina de Turing. Entonces, el enunciado

" T_n se para"

es una proposición —escribámosla *S*(*n*)— que podemos expresar en nuestro sistema formal, para cada número natural *n*.

La proposición *S*(*n*) será verdadera para algunos valores de *n*, y falsa para otros. El conjunto de *todas* las *S*(*n*), cuando *n* recorre los números naturales 0, 1, 2, 3,... estará representado como un subconjunto *S* de $\mathbb{N}$. Recuérdese ahora el resultado fundamental de Turing (capítulo II,) de que no hay algoritmo que afirme que "$T_n(n)$ no se para" precisamente en esos casos en los que efectivamente $T_n(n)$ no se para. Esto prueba que el conjunto de las *S*(*n*) *falsas* no es recursivamente numerable.

Observemos que la parte de S que está en P consta de aquellas $S(n)$ que son *verdaderas*. ¿Por qué es así? Ciertamente, si una $S(n)$ particular es demostrable, entonces debe ser verdadera (debido a que hemos escogido un sistema formal "razonable"), así que la parte de S que está en P debe constar solamente de proposiciones $S(n)$ *verdaderas*. Más aún, ninguna proposición $S(n)$ verdadera puede estar fuera de P, porque si $T_n(n)$ se detiene podemos demostrar, dentro del sistema, que efectivamente lo hace.*

Imaginemos a continuación que el complemento de P es recursivamente numerable. Entonces tendremos algún algoritmo con el cual generar los elementos del conjunto complementario. Podríamos poner en marcha ese algoritmo y anotar las proposiciones $S(n)$ que encontremos. Todas ellas serían las $S(n)$ falsas, así que nuestro método nos proporcionaría una enumeración recursiva del conjunto de $S(n)$ falsas. Pero como señalábamos arriba, las $S(n)$ falsas *no* son recursivamente numerables. Tal contradicción establece que, después de todo, el complemento de P no puede ser enumerado recursivamente. Por consiguiente, el *conjunto P no es recursivo*, que es lo que buscábamos dejar establecido.

Estas propiedades prueban que nuestro sistema formal no puede ser completo: debe haber proposiciones que ni son demostrables ni refutables dentro de él. Sin esas proposiciones "indecidibles", el complemento del conjunto P tendría que ser el conjunto de proposiciones *refutables* (todo lo que no es demostrable es refutable). Pero hemos visto que las proposiciones refutables constituyen un conjunto recursivamente numerable, de modo que P sería *recursivo*. Sin embargo, P *no* es recursivo —una contradicción que establece la requerida característica de incompletitud—. Esta es la principal estocada del teorema de Gödel.

¿Qué sucede entonces con el subconjunto T de $\mathbb{N}$ que representa las proposiciones *verdaderas* de nuestro sistema formal? ¿T es recursivo?, ¿es recursivamente numerable? ¿Lo es el complemento de T? La respuesta a todas estas preguntas es "no". Observemos que las proposiciones falsas de la forma

"$T_n(n)$ se para"

no pueden ser generadas mediante un algoritmo y, en consecuencia, las proposiciones falsas en *conjunto* no pueden ser generadas mediante un algoritmo, puesto que semejante algoritmo debería enumerar, en particular, todas las anteriores proposiciones " $T_n(n)$ se para" falsas.

Igual, el conjunto de las proposiciones *verdaderas* no puede ser generado mediante un algoritmo (ya que semejante algoritmo podría ser modificado de forma trivial para producir todas las proposiciones falsas sin más que hacerle tomar la *negación* de cada proposición que genera). Puesto que las proposiciones verdaderas (y las falsas) no son recursivamente numerables, aquellas constituyen una colección más complicada y profunda que la de las proposiciones demostrables dentro del sistema.

Esto ilustra, una vez más, un aspecto del teorema de Gödel: que el concepto de *verdad* matemática es sólo parcialmente accesible a través del argumento formal.

Existen, no obstante, algunas clases de proposiciones aritméticas que forman conjuntos recursivamente enumerables. Por ejemplo, las proposiciones verdaderas de la forma

* La demostración constaría, de hecho, de una serie de pasos que reflejaran la acción de la máquina funcionando hasta que se pare. La demostración será completa una vez que la máquina se detenga.

$$\exists w,x,...,z\ [f(w,x,...,z)=0],$$

donde *f*( ) es una función construida a partir de las operaciones aritméticas ordinarias de adición, sustracción, multiplicación, división y elevación a una potencia, constituyen un conjunto recursivamente numerable (que denominaré A), como no es difícil de ver.[8] Un ejemplo de una proposición de esta forma — aunque no sabemos si es cierta — es la negación del "último teorema de Fermat", para lo que podemos tomar *f*( ) dada por

$$f(w, x, y, z) = (x + 1)^{w+3} + (y + 1)^{w+3} - (z + 1)^{w+3}$$

Sin embargo, el conjunto *A* resulta ser no recursivo (un hecho que *no* es tan fácil de ver — aunque es una consecuencia del auténtico argumento original de Gödel). Por lo tanto, no disponemos de ningún algoritmo mediante el cual podamos, ni en principio, conocer la verdad o falsedad del "último teorema de Fermat".

En la fig. IV. 1 está representado un conjunto recursivo como una región con un contorno sencillo, de modo que se puede saber directamente si un punto pertenece o no al conjunto. Podemos pensar que cada punto de la figura representa un número natural. El conjunto complementario está entonces representado también como una región de apariencia sencilla. En la fig. IV.2 he tratado de representar un conjunto recursivamente numerable pero *no* recursivo como un conjunto con un contorno complicado, en donde se supone que el conjunto a un lado del contorno — el lado recursivamente enumerable — parece más sencillo que el que está al otro lado. Las figuras son muy esquemáticas, y no se pretende que sean "geométricamente precisas" en ningún sentido. En particular, no tiene especial significación el hecho de que esas figuras hayan sido representadas en un plano bidimensional.



FIGURA IV. 1. *Representación muy esquemática de un conjunto recursivo.*

[8] Enumeramos los conjuntos {*v*, *w*, *x*,..., *z*} donde *v* representa la función *f* según algún esquema lexicográfico. Hacemos una comprobación (recursivamente) en cada paso para ver si $f(w, x,..., z) = 0$ y retenemos la proposición $\exists w,x, ..., z\ [f(w,x,..., z) = 0]$ sólo en caso de que sea así.



FIGURA IV.2. *Representación muy esquemática de un conjunto recursivamente numerable —región negra— que no es recursivo. La idea consiste en que la región blanca está definida sólo como "lo que queda" cuando se elimina la región negra generada computablemente y no es computable el hecho de que un punto esté en la región blanca.*

En la fig. IV.3 he sugerido la manera en que las regiones *P*, *T* y *A* yacen dentro del conjunto ℕ.

¿ES RECURSIVO EL CONJUNTO DE MANDELBROT?

Los conjuntos no recursivos tienen la propiedad de ser esencialmente complicados. Su complejidad desafía todos los intentos de sistematización, porque de no ser así, esa misma sistematización conduciría hacia algún método algorítmico apropiado. En el caso de un conjunto no recursivo, no hay una vía algorítmica general que sirva para saber si un elemento (o "punto") pertenece o no al conjunto. Ahora bien, al comienzo del capítulo III fuimos testigos de un cierto conjunto de apariencia extraordinariamente complicada: el conjunto de Mandelbrot. Aunque las reglas que proporcionan su definición son sorprendentemente simples, el propio conjunto exhibe una variedad sin fin de estructuras altamente elaboradas. ¿Podría ser éste un ejemplo de conjunto no recursivo?

Este complicado paradigma ha sido conjurado, para que nuestros ojos puedan verlo, mediante la "magia" de la moderna tecnología de la computación electrónica. ¿No es ésta, acaso, la acción algorítmica encarnada? Debemos considerar, empero, de qué forma la computadora produce realmente estas imágenes.

Para verificar si un punto del plano de Argand —un número complejo *c*— pertenece al conjunto de Mandelbrot (en color negro) o al conjunto complementario (en color blanco), la computadora empieza con O, luego aplica la iteración

$$z \to z^2 + c$$

a $z = 0$ para obtener c, y luego a $z = c$ para obtener $c^2 + c$ y después a $z = c^2 + c$ para obtener $c^4 + 2c^3 + c^2 + c$, y así sucesivamente. Si esta secuencia $0, c, c^2 + c, c^4 + 2c^3 + c^2 + c, \ldots$ permanece acotada, entonces el punto representado por c se colorea en negro; en caso contrario, se colorea en

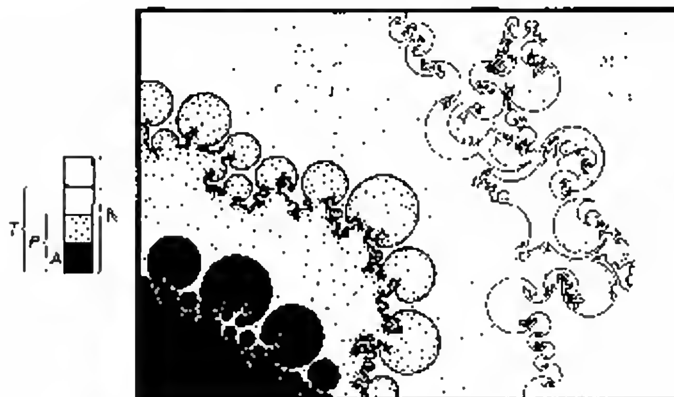


FIGURA IV.3. *Representación muy esquemática de varios conjuntos de proposiciones. El conjunto* P *de las proposiciones que son demostrables dentro del sistema es, como el* A, *recursivamente numerable pero no recursivo*; *el conjunto* T *de las proposiciones verdaderas no es siquiera recursivamente enumerable.*

blanco. ¿Cómo determina la máquina si la secuencia permanece o no acotada? En principio, esta pregunta implica saber qué sucede tras un *infinito* número de términos de la secuencia. Esto no es materia computable, pero hay formas de determinar, tras un número finito de términos, cuándo la secuencia se ha convertido en no acotada. (De hecho, una vez que alcanza el círculo de radio $1 + \sqrt{2}$ centrado en el origen podemos estar seguros de que la secuencia es no acotada.)

Así, en cierto sentido, el complemento del conjunto de Mandelbrot (es decir, la región *blanca*) es recursivamente numerable. Si el número complejo *c* está en la región blanca, existe un algoritmo para verificarlo. ¿Qué sucede con el propio conjunto de Mandelbrot, es decir la región negra? ¿Existe un algoritmo que nos asegure que un punto sospechoso de estar en la región negra está efectivamente ahí? Por el momento no se conoce la respuesta a esta pregunta.[9] He consultado a varios colegas y expertos, y ninguno parece saber de tal algoritmo, pero tampoco han tropezado con alguna demostración de que no exista. Al menos, no parece haber ningún algoritmo *conocido* para la región negra. Tal vez el complemento del conjunto de Mandelbrot es en verdad un conjunto recursivamente numerable que no es recursivo.

Antes de explorar más a fondo esta sugerencia, abordemos otros temas útiles para nuestros conocimientos de la computabilidad en física.

He sido algo inexacto en la discusión precedente. He aplicado términos como "recursivamente numerable" y "recursivo" a conjuntos de puntos en el plano de Argand, o sea a conjuntos de números complejos. Estos términos deberían utilizarse estrictamente sólo para los números naturales y otros conjuntos *numerables.*

Hemos visto en el capítulo III , que los números reales no son numerables y, por lo tanto, tampoco los números complejos, ya que los números reales pueden ser considerados como tipos particulares de números complejos, digamos números complejos con partes imaginarias que se desvanecen . De hecho, hay "tantos" números complejos como números reales, a saber, "*C*" de ellos. Para establecer una relación aproximada uno-a-uno entre los números complejos y los

[9] Recientemente, Leonore Blum me comentó, después de leer la edición en pasta dura de este libro, que había comprobado que el complemento del conjunto de Mandelbrot, en efecto, es no recursivo, tal como lo aventuro en el texto, en el particular sentido al que se refiere la nota 10.

números reales, podemos tomar las expansiones decimales de las partes real e imaginaria de cada número complejo e intercalarlas para dar los dígitos (impares y pares) del correspondiente número real: *v.g.* el número complejo 3.6781... + i 512.975... correspondería al número real 50132.6977851....

Un modo de evitar este problema sería referirnos sólo a números complejos *computables*, pues vimos en el capítulo III que los números reales computables — y por lo tanto, también los números complejos computables, — son numerables. Sin embargo, tropezamos con una dificultad: no existe ningún algoritmo general que establezca si dos números computables, dados en términos de sus respectivos algoritmos, son iguales entre sí o no. (Podemos determinar algorítmicamente su diferencia, pero no podemos saber algorítmicamente si tal diferencia es igual a cero. Imaginemos dos algoritmos que generan los dígitos 0.99999... y 1.00000..., respectivamente: jamás sabremos si los 9's o los 0's continuarán de modo indefinido, por lo cual los números serán siempre iguales, o si finalmente aparecerá algún otro dígito, y los números serán desiguales.)

Por consiguiente, nunca podríamos saber si tales números son iguales. Una de las implicaciones de esto es que incluso con un conjunto tan sencillo como el *disco unidad* en el plano de Argand (el conjunto de puntos cuya distancia al origen no es mayor que la unidad, es decir, la región negra en la fig. IV.4), no habría algoritmo que estableciera con certeza si un número complejo yace en el disco o no. El problema no se presenta con los puntos en el interior (o con los puntos fuera del disco), sino sólo con puntos que yacen sobre el mismo borde del disco, es decir, en el círculo unidad en sí. El círculo unidad se considera parte del disco. Si se nos diera simplemente un algoritmo *capaz* de generar los dígitos de las partes real e imaginaria de algún número complejo y si existiese la duda acerca de si este número complejo yace sobre el círculo unidad, no necesariamente podríamos verificarlo. No hay algoritmo *capaz* de establecer si el número computable

$$x^2+y^2$$

es igual a 1 o no, siendo éste el criterio para determinar si el número complejo computable $x + iy$ yace o no sobre el círculo unidad.

No es esto lo que buscamos. El disco unidad *debería* considerarse entre los recursivos. No hay muchos conjuntos más sencillos que el disco unidad.

Podríamos *soslayar* la existencia del contorno. Para puntos que están estrictamente en el interior o estrictamente en el exterior, existen algoritmos para verificarlo. (No hay más que generar los dígitos de $x^2 + y^2$ uno tras otro, y encontrar finalmente un dígito diferente de 9 tras el punto decimal en 0.99999... o diferente de 0 en 1.00000...) En tal sentido, el disco unidad *es* recursivo, pero este punto de vista es incómodo porque matemáticamente necesitamos expresar argumentos en términos de lo que sucede *en* el contorno. Es posible, por el contrario, que semejante punto de vista fuera apropiado para la física. Sobre esto volveremos más adelante.

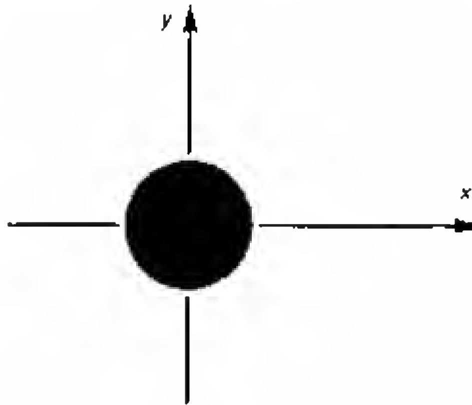


FIGURA IV.4. *El disco unidad contaría ciertamente entre los conjuntos "recursivos", pero para eso requeriría un punto de vista particular.*

Podríamos adoptar un enfoque que no haga referencia a números complejos computables. En lugar de enumerar los números complejos en el interior o el exterior del conjunto en cuestión, podemos buscar un algoritmo que establezca, *dado* el número complejo, si yace en el conjunto o si yace en el complemento del conjunto. Por "dado" quiero decir que para cada número complejo que estamos verificando, se nos presentan uno tras otro los sucesivos dígitos de las partes real e imaginaria.

No exijo que exista un algoritmo, conocido o desconocido, para *presentar* estos dígitos. Un conjunto de números complejos se consideraría "recursivamente enumerable" siempre que existiese un solo algoritmo tal, que siempre que se le presentase una sucesión de dígitos semejante respondiera finalmente "sí", tras un número finito de pasos, si y sólo si el número complejo yace de verdad en el conjunto. Al igual que sucedía con el primer punto de vista sugerido, este nuevo enfoque "omite" los contornos. Tanto el interior como el exterior del disco unidad serán tomados como numerablemente recursivos, mientras que el propio contorno no lo haría.

No está completamente claro para mí que cualquiera de estos puntos de vista sea realmente el que necesitamos.[10] Cuando se aplica al conjunto de Mandelbrot, la filosofía de "omitir el contorno" puede dejar de lado mucha de la complejidad del conjunto. Este conjunto consta en parte de "gotas" — regiones con interiores — y en parte de "zarcillos". La complicación más extrema parece estar en los zarcillos, que pueden retorcerse violentamente. Sin embargo, éstos no están en el interior del conjunto y únicamente serían "omitidos" si adoptáramos una de las dos filosofías.

Pese a todo, todavía no está claro que el conjunto de Mandelbrot sea "recursivo", aun cuando sólo se consideren las gotas. La cuestión parece descansar en cierta conjetura relacionada con el conjunto de Mandelbrot: ¿es éste lo que se denomina "localmente conexo"? No me propongo explicar aquí el significado o la importancia del término, sino indicar que se trata de temas

---

[10] Existe una nueva teoría de la computabilidad para funciones reales de números reales (en contraste con las convencionales funciones que toman valores naturales para números naturales), debida a Blum, Shub y Smale (1989), cuyos detalles sólo muy recientemente han llamado mi atención. Esta teoría se aplicaría también a funciones complejas y podría tener una relación importante con algunos de los temas planteados en el texto.

difíciles que plantean preguntas no resueltas en torno al conjuntos de Mandelbrot, y algunas de ellas inclusive en la vanguardia de la investigación matemática.

Hay otros puntos de vista para no problematizar el hecho de que los números complejos no son numerables. En lugar de considerar *todos* los números complejos computables, podemos considerar un subconjunto de números con los cuales una computadora pueda determinar si dos de ellos son o no iguales. Un subconjunto sencillo sería el de los números complejos "racionales" en el que las partes real e imaginaria de los números son, ambas, números racionales. No creo que esto eliminara muchos de los "zarcillos" del conjunto de Mandelbrot, aun cuando este punto de vista es restrictivo. Más satisfactorio sería considerar los números *algebraicos*, aquellos números complejos que son soluciones de ecuaciones algebraicas con coeficientes enteros. Por ejemplo, todas las soluciones de

$$129 z^7 - 33 z^5 + 725 z^4 + 16 z^3 - 3 z - 3 = 0$$

son números algebraicos. Los números algebraicos son numerables y computables, y es ciertamente una cuestión de computabilidad el decidir si dos de ellos son o no iguales. (Resulta que muchos de ellos yacen sobre el contorno del círculo unidad y en los zarcillos del conjunto de Mandelbrot.) Podemos plantear en esos términos la cuestión de si el conjunto de Mandelbrot es o no es recursivo.

Tal vez los números algebraicos sean apropiados en el caso de los dos conjuntos recién considerados, pero en general no resuelven nuestras dificultades. Consideremos el conjunto (la región negra de la fig. IV.5) definido por la relación

$$y \geq e^x$$

para $x + iy$ (= z) en el plano de Argand.

Tanto el interior del conjunto como el interior del complemento del conjunto son recursivamente numerables, con arreglo a cualquiera de los puntos de vista expresados más arriba. Pero (como se deduce de un famoso teorema demostrado en 1882 por F. Lindemann) el contorno, $y = e^x$, contiene sólo *un* punto algebraico, a saber, el punto z = i. Los números algebraicos no nos ayudan aquí para explorar la naturaleza algorítmica del contorno. No sería difícil encontrar otra subclase de números computables que bastaran en este caso particular. Nos quedamos, empero, con la sensación de que todavía no ha sido alcanzado el punto de vista correcto.

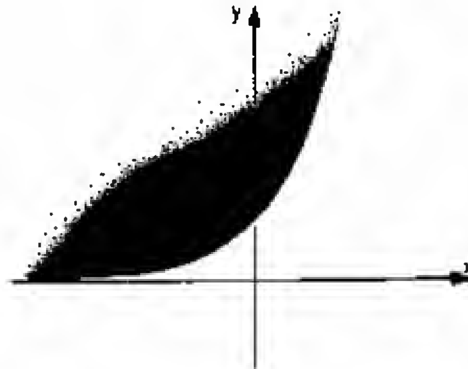


FIGURA IV.5. *El conjunto definido por la relación exponencial* $y \geq e^x$ *debería estar también entre los "recursivos".*

ALGUNOS EJEMPLOS DE MATEMÁTICAS NO RECURSIVAS

Existen muchas áreas de las matemáticas en las que aparecen problemas que son no recursivos. Así, se nos puede presentar una clase de problemas para los que la respuesta en cada caso es o "si" o "no", pero para los que no existe ningún algoritmo general que determine cuál de los dos es el caso. Algunas de estas clases de problemas tienen una apariencia notablemente sencilla.

Consideremos, en primer lugar, el problema de encontrar soluciones enteras de sistemas de ecuaciones algebraicas con coeficientes también enteros. Tales ecuaciones se llaman ecuaciones *diofánticas* (en honor del matemático griego Diofanto, que vivió en el siglo III antes de Cristo y estudió este tipo de ecuaciones). Un conjunto de tales ecuaciones podría ser

$$z^3 - y - 1 = 0 \qquad yz^2 - 2x - 2 = 0, \qquad y^2 - 2xz = 0$$

y el problema consiste en determinar si tienen solución para valores *enteros* de x, y, z. En este caso particular sí la tienen, y la solución está dada por

$$x=13, \qquad y = 7, \qquad z = 2$$

Sin embargo, no existe algoritmo que establezca lo mismo en el caso de un conjunto arbitrario** de ecuaciones diofánticas. La aritmética diofántica, pese a la naturaleza elemental de sus ingredientes, es parte de las matemáticas no algorítmicas.

(Un ejemplo algo menos elemental lo constituye la *equivalencia topológica de variedades*. Lo menciono debido a su importancia para los temas discutidos en el capítulo VIII. Para entender lo que es una "variedad" consideremos primero un lazo de cuerda, que es una variedad de *una* sola dimensión, y consideremos luego una superficie cerrada, una variedad de *dos* dimensiones. A continuación tratemos de imaginar una "superficie" con *tres* o un número mayor de dimensiones. "Equivalencia topológica" de dos variedades significa que una de ellas puede ser deformada hasta que coincida con la otra en un movimiento continuo, sin raspar ni pegar. Por ejemplo, una superficie esférica y la superficie de un cubo son topológicamente equivalentes, mientras que ambas son topológicamente no equivalentes a la superficie de un anillo o de una taza de té, siendo estas dos últimas topológicamente equivalentes entre sí.

Ahora bien, para variedades bidimensionales existe un algoritmo útil para saber si dos de ellas son o no topológicamente equivalentes — lo que corresponde, en realidad, a contar el número de "asas" que tiene cada superficie. En el caso de tres dimensiones todavía no se conoce una respuesta, pero sí se sabe que para cuatro o más dimensiones *no* existe algoritmo que ayude a conocer la equivalencia. El caso tetradimensional tiene cierta importancia para la física a partir de que, según la teoría de la relatividad general de Einstein, espacio y tiempo constituyen conjuntamente una 4-variedad; véase capítulo V. Geroch y Hartle 1986 han sugerido que esta propiedad no algorítmica podría tener importancia para la "gravitación cuántica"; cfr. también capítulo VIII.)

Consideremos un problema distinto: el *problema de las palabras*.[11] Supongamos que tenemos algún alfabeto de símbolos y consideremos varias cadenas de estos símbolos, que llamaremos

---

[*] Esto da una respuesta negativa al décimo problema de Hilbert mencionado (Véase, por ejemplo, Devlin, 1988.) Aquí el número de variables no está restringido, aunque se sabe que no son necesarias más de nueve para que la propiedad no-algorítmica sea válida.

[11] Este problema particular se denomina con más propiedad "el problema de las palabras para semigrupos". Existen también otras formas del problema de las palabras en las que las reglas son ligeramente diferentes No nos ocuparemos de esto aquí.

palabras. Las palabras mismas no necesitan tener un significado, pero habrá una lista (finita) de "igualdades" entre ellas, que usaremos para derivar otras nuevas "igualdades". Esto se hará sustituyendo palabras de la lista inicial con otras palabras (normalmente más largas) que las contengan como porciones. Cada una de esas porciones puede ser reemplazada por otra porción que, según la lista, se considere igual a ella. El problema entonces es decidir, para un par de palabras dado, si de acuerdo con estas reglas son "iguales" o no.

Como ejemplo podríamos tener en nuestra lista inicial

LAS = AS

ASO = A

NASO = RON

SAN = LIRÓN

GAS = DEL

A partir de éstas podemos derivar, por ejemplo

BOA = BOLA

mediante el uso de sustituciones sucesivas de la segunda, la primera, y de nuevo la segunda de las relaciones de la lista inicial:

BOA = BOASO = BOLASO = BOLA

El problema es: dado un par de palabras, ¿podemos obtener una a partir de la otra utilizando simplemente estas sustituciones? ¿Podemos, por ejemplo, ir de GASOLINA a DEAN o, pongamos por caso, de GASTAR a DELATAR? En el primer caso la respuesta resulta ser "sí", mientras que en el segundo caso es "no". Cuando la respuesta es "sí", la manera normal de probarlo sería simplemente mostrando una cadena de igualdades en la que cada palabra se obtiene de la precedente mediante el uso de una relación permitida. Así (indicando en **negritas** las letras que van a ser cambiadas, y en *itálicas* las letras que acabamos de cambiar):

GA**SO**LINA = CALIN**A** = GALI**N**ASO — GA**LI**RON = **GA**SAN

= *DELAN* = DELASON = DEASON = DEAN

Cómo podemos decir que es imposible ir de GASTAR a DELATAR por medio de las reglas permitidas? Necesitamos pensar un poco más, pero no es difícil ver que existen varias formas de lograrlo. La más simple puede ser la siguiente: en cada "igualdad" de nuestra lista inicial, el número de As más el número de Rs más el número de Ds es el mismo en cada lado. Por lo tanto, el número total de As, Rs y Ds no puede cambiar a lo largo de cualquier sucesión de sustituciones permitidas. Pero como dicho número es 3 para GASTAR mientras que es 4 para DELATAR, no hay sustituciones permitidas para ir de GASTAR a DELATAR.

Nótese que cuando las dos palabras son "iguales" podemos probarlo exhibiendo una cadena formal de símbolos permitidos, con las reglas que nos han sido dadas, mientras que en el caso en que son "desiguales" tenemos que recurrir a argumentos *acerca* de esas reglas. Hay un algoritmo evidente que podemos utilizar para establecer la "igualdad" entre palabras, siempre que las palabras *sean* efectivamente "iguales". Todo lo que tenemos que hacer es un listado lexicográfico

de todas las secuencias posibles de palabras, y luego tachar de la lista las cadenas en las que haya un par de palabras consecutivas donde la segunda no se siga de la primera siguiendo las reglas establecidas. Las secuencias restantes proporcionan todas las "igualdades" buscadas entre palabras. Sin embargo, no existe, en general, un algoritmo tan obvio para determinar cuándo dos palabras *no* son "iguales", y tenemos que recurrir a la "inteligencia" para establecer el hecho. (Realmente necesité algún tiempo antes de encontrar un "truco" para establecer que GASTAR y DELATAR *no* son "iguales"; con otro ejemplo podría ser necesario un "truco" diferente. La inteligencia, dicho sea de paso, es también útil —aunque no indispensable— para establecer la *existencia* de una "igualdad".)

No es excesivamente difícil encontrar un algoritmo que verifique si dos palabras de la lista particular dada al principio son "desiguales" cuando efectivamente lo son. Y no obstante, para *encontrar* el algoritmo que funcione en este caso necesitamos ejercitar la inteligencia. Porque no hay un único algoritmo que se pueda utilizar universalmente para *todas* las elecciones posibles de la lista inicial; no hay solución algorítmica para el problema de las palabras. El problema general de la palabra pertenece a las matemáticas no recursivas.

Existen incluso ciertas selecciones *particulares* de la lista inicial para las que no hay ningún algoritmo con el cual saber cuándo dos palabras son desiguales. Una de éstas viene dada por

AH = HA

OH = HO

AT = TA

OT = TO

TAI = IT

HOI = IH

THAT = ITHT

(Esta lista está adaptada de la presentada en 1955 por G. S. Tseitin y Dana Scott; véase Gardner, 1958, p. 144.)

Por lo tanto, el problema particular de las palabras constituye por *si mismo* un ejemplo de matemáticas no recursivas, en el sentido de que si usamos esta lista inicial particular no podemos decidir algorítmicamente si dos palabras dadas son o no "iguales".

El problema general de las palabras surgió a partir de consideraciones de lógica matemática formalizada ("sistemas formales", etc.). La lista inicial juega el papel de un sistema de axiomas, y la regla de sustitución de palabras, el papel de las reglas formales de inferencia. La demostración de la no recursividad del problema de la palabra surge de ahí.

Como ejemplo final de un problema matemático que es no recursivo, consideremos la cuestión del recubrimiento del plano euclidiano con formas poligonales, en donde se nos da un número finito de formas diferentes y se pregunta si es posible recubrir completamente el plano, sin huecos ni solapamientos, con el mero empleo de estas formas. Una disposición de formas semejante se denomina una *teselación* del plano. Estamos familiarizados con el hecho de que estas teselaciones son posibles si sólo se utilizan cuadrados, o sólo triángulos equiláteros, o sólo

hexágonos regulares (como se ilustra en la fig. X.2, capítulo X), pero no si sólo se utilizan pentágonos regulares.

Habrá muchas otras formas que teselarán el plano, tales como cada uno de los dos pentágonos *irregulares* ilustrados en la fig. IV.6. Con un *par* de formas las teselaciones pueden ser mucho más elaboradas. En la fig. IV.7 se dan dos ejemplos sencillos. Todos los ejemplos dados hasta ahora tienen la propiedad de ser *periódicos*, lo que significa que son exactamente repetitivos en dos direcciones independientes. En términos matemáticos, decimos que existe un *paralelogramo periodo* —un paralelogramo tal que, si lo marcamos de alguna manera y luego lo repetimos una y otra vez en las dos direcciones paralelas a sus lados, reproducirá el patrón de la teselación dada. En la fig. IV.8 se muestra un ejemplo en el que una teselación periódica con una tesela en forma de espina es representada a la izquierda y relacionada con un paralelogramo periodo cuya teselación periódica se muestra a la derecha.

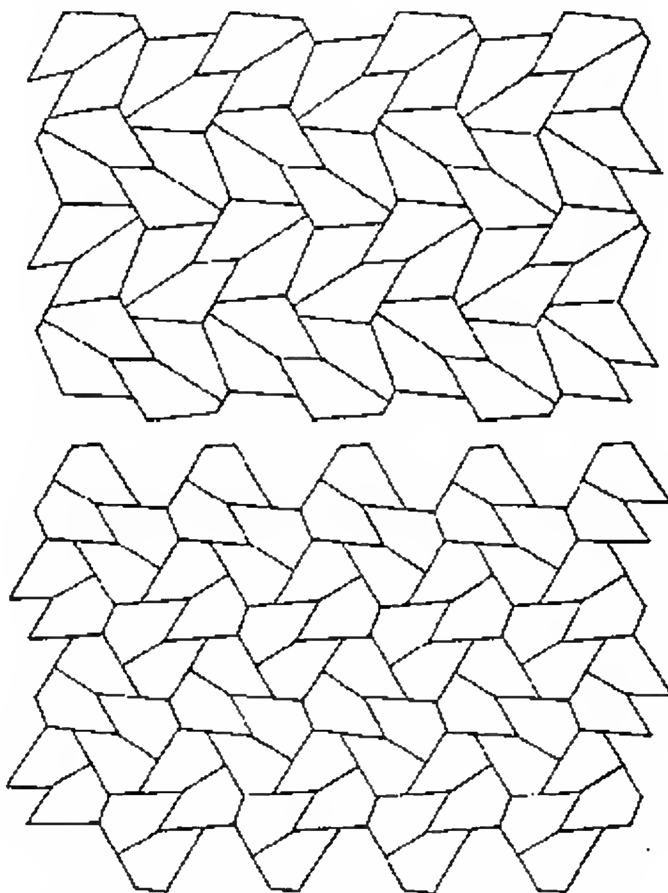


FIGURA IV.6. *Dos ejemplos de teselaciones periódicas del plano que utilizan —cada una— una sola forma, descubiertos por Marjorie Rice en 1976.*

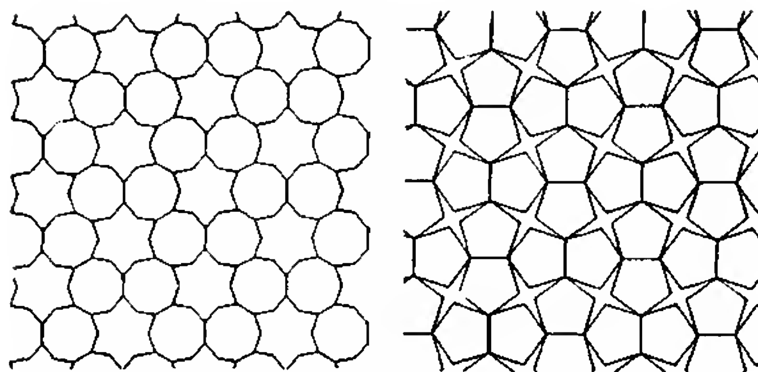


FIGURA IV. 7. *Dos ejemplos de teselaciones periódicas del plano, de dos formas cada una.*

Ahora bien, existen muchas teselaciones del plano que *no* son periódicas. La fig. IV.9 muestra tres teselaciones "espirales" no periódicas con la misma tesela en forma de espina que la de la fig. IV.8. Esta forma particular de tesela se conoce como *versatile*[*] y fue ideada por B. Grünbaum y G. C. Shephard (1981, 1987), aparentemente con base en una forma anterior debida a H. Voderberg. Nótese que el versátil teselará tanto periódica como no periódicamente. Esta propiedad es compartida por muchas otras formas de tesela y conjuntos de formas de tesela. ¿Existen teselas simples o conjuntos de ellas que teselen el plano *sólo* no periódicamente? La respuesta a esta pregunta es "sí". En la fig. IV. 10 he representado un conjunto de seis teselas, construido por el matemático estadounidense Raphael Robinson (1971), que teselan el plano entero pero sólo de manera no periódica.

Merece la pena entrar un poco en la historia de cómo surgió este conjunto no periódico de teselas (*cfr*. Grünbaum y Shepard, 1987). En 1961 el lógico chino-estadounidense Hao Wang abordó la cuestión de si hay o no un *procedimiento de decisión* para el problema de la teselación, es decir, si existe un *algoritmo* para prever si un conjunto finito de formas poligonales diferentes teselará el plano entero.[**] Wang demostró que tal procedimiento de decisión existiría si se pudiera probar que todo conjunto finito de teselas diferentes que tésele el plano, también lo hará periódicamente.

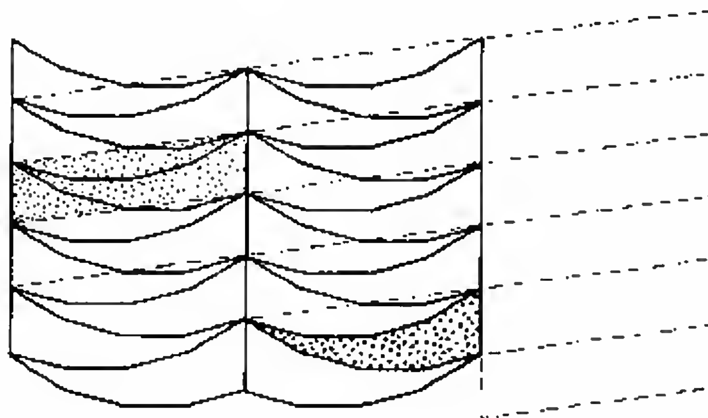


FIGURA IV. 8. *Una teselación periódica, ilustrada en relación con su paralelogramo periodo.*

[*] Juego de palabras entre "tile" = tesela y "versatile" = versátil. [N. del T.]

[**] En realidad, Hao Wang concibió un problema ligeramente distinto —con teselas cuadradas, sin rotación y con bordes parejos coloreados, pero estas diferencias no son importantes aquí para nosotros.

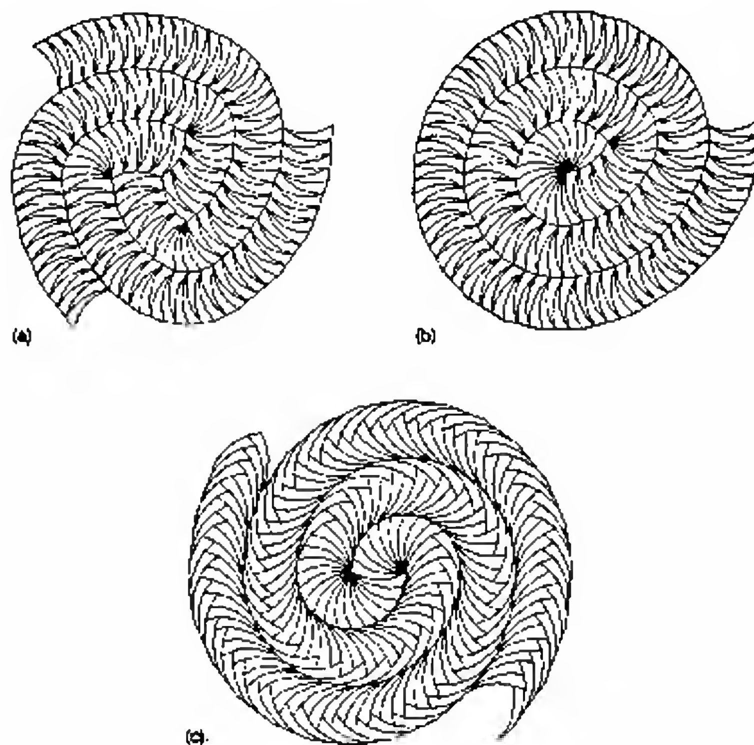


FIGURA IV.9. *Tres teselaciones "espirales" no periódicas que utilizan la misma forma "versátil" que se usó en la fig. IV.8.*

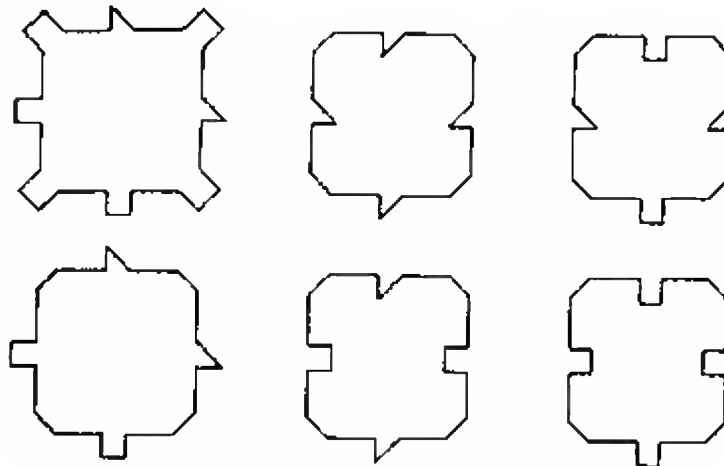


FIGURA IV. 10. *Las seis teselas de Raphael Robinson que teselan el plano sólo de forma no periódica.*

Yo pienso que en esa época existía la sensación generalizada de que era poco probable que pudiera existir un conjunto que violase esta condición, es decir, un conjunto "aperiódico" de teselas. Sin embargo, en 1966, Robert Berger fue capaz de probar, siguiendo algunas de las líneas que Hao Wang había sugerido, que *no* existe procedimiento de decisión para el problema

de la teselación: el problema de la teselación forma también parte de las matemáticas no recursivas.[12]

Del resultado previo de Hao Wang se deduce que debe existir un conjunto aperiódico de teselas, y Berger fue capaz de probarlo, encontrando el primero de estos conjuntos. Pero, debido a lo complicado de esta línea argumental, su conjunto incluía un número desorbitadamente alto de teselas diferentes, originalmente 20426. Con algún ingenio adicional, Berger pudo reducir su número a 104, y más tarde, en 1971, Raphael Robinson logró hacerlo hasta las seis mostradas en la fig. IV. 10.

En la figura IV. 11 se muestra otro conjunto aperiódico de seis teselas. Yo mismo lo presenté en 1973, siguiendo una línea de pensamiento bastante diferente. (Volveré a este tema en el capítulo X en cuya fig. X.3 se representa una mosaico teselado con estas formas.) Después de que el conjunto aperiódico de seis teselas de Robinson hubiera llamado mi atención, pensé varias operaciones de recortado y repegado y con ellas pude reducirlo a dos. En la fig. IV. 12 se muestran dos esquemas opcionales. Las estructuras necesariamente no periódicas que exhiben las teselaciones completas tienen muchas propiedades notables, incluso una estructura cuasi-periódica con simetría quíntuple aparentemente imposible en cristalografía. Más tarde volveremos sobre este punto.

Es notable que una área de las matemáticas tan "trivial" como ésta —a saber, el recubrimiento del plano con formas congruentes—, que parece casi un juego de niños, forme parte de las matemáticas no recursivas, pero en esa área hay muchos problemas difíciles y no resueltos. No se sabe, por ejemplo, si existe un conjunto aperiódico que conste de *una sola* tesela.

El problema de la teselación, tal como lo tratan Wang, Berger y Robinson, utiliza teselas basadas en cuadrados. Yo admito aquí polígonos de forma general, y se requiere un método de computación para mostrar las teselas individuales. Una manera de hacerlo sería señalar sus vértices como puntos en el plano de Argand, y estos puntos pueden perfectamente darse como números algebraicos.

¿ES EL CONJUNTO DE MANDELBROT SEMEJANTE A LA MATEMÁTICA NO RECURSIVA?

Volvamos ahora a nuestro anterior examen sobre el conjunto de Mandelbrot. Voy a suponer, con propósito de ilustración, que el conjunto de Mandelbrot es no recursivo en un sentido. Puesto que su complemento es recursivamente numerable, esto querría decir que el propio conjunto no sería recursivamente numerable. Es posible que la forma del conjunto de Mandelbrot nos dé algunas lecciones acerca de la naturaleza de los conjuntos no recursivos y las matemáticas no recursivas. Volvamos a la Fig. III.2 que ya encontramos en el capítulo III.

---

[12] Hanf (1974) y Myers (1974) han demostrado, además, que hay un solo conjunto (de un gran número de teselas) que teselará el plano sólo de un modo *no-computable*.

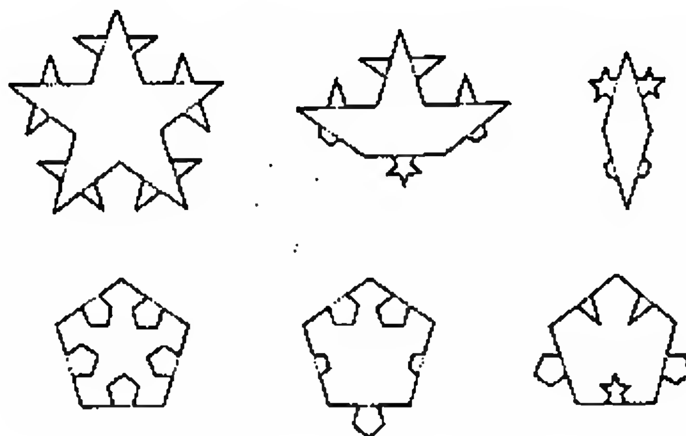


FIGURA IV. 11. *Otro conjunto de seis teselas que tesela el plano sólo de forma no periódica.*

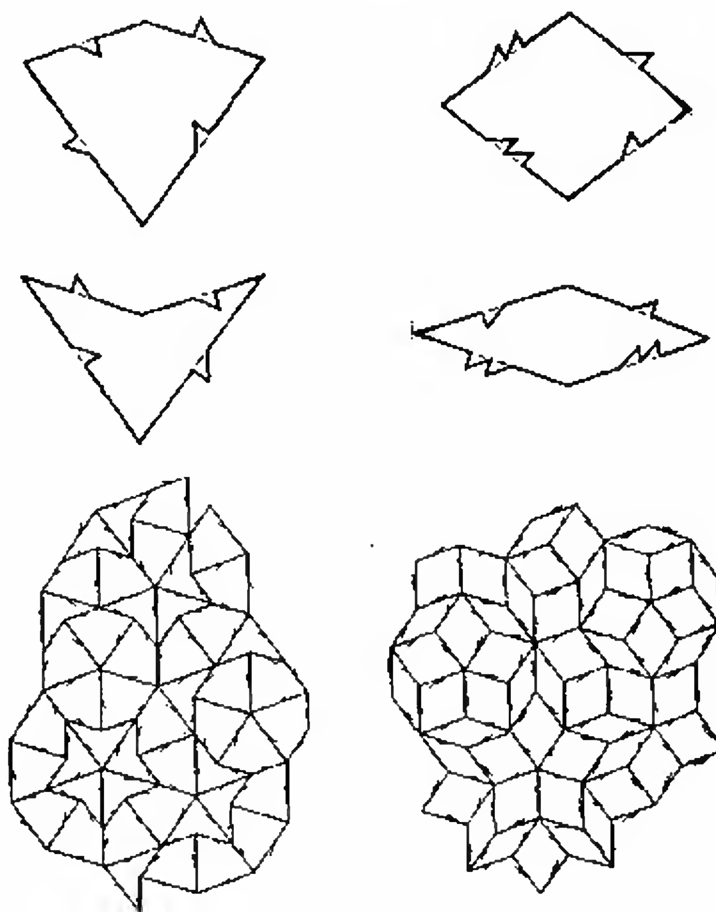


FIGURA IV. 12. *Dos pares, cada uno de los cuales teselará sólo de forma no periódica, "teselas de Penrose", y las regiones del plano teseladas con cada par.*

La mayor parte del conjunto parece estar ocupada por una extensa región en forma de corazón, que he llamado A en la fig. IV. 13. La forma se conoce como un *cardioide* y su región interior se define matemáticamente como el conjunto de puntos *c* del plano de Argand que surgen de la expresión

$$c = z - z^2,$$

donde z es un número complejo cuya distancia al origen es menor que ½. Este conjunto es recursivamente enumerable en el sentido sugerido antes: existe un algoritmo tal que, cuando se aplica a un punto en el interior de la región, verificará que el punto está efectivamente en dicha región interior. El algoritmo real se obtiene a partir de esa fórmula.

Consideremos ahora la región en forma de disco inmediatamente a la izquierda del cardioide principal (región B en la fig. IV.13). Su interior es el conjunto de puntos

$$c = z - 1$$

donde z dista del origen menos de 1/4 .

Esta región es el interior de un disco. Es decir, se trata del conjunto de puntos dentro de un círculo exacto. De nuevo esta región es recursivamente enumerable en el sentido anterior. ¿Qué sucede con las otras "verrugas" del cardioide? Consideremos las dos verrugas que siguen en tamaño. Estas son las gotas más o menos circulares que aparecen encima y debajo del cardioide en la fig. III.2 y que están marcadas como C_1, C_2 en la fig. IV. 13. Vienen dadas en términos del conjunto

$$c^3 + 2c^2 + (1-z)c + (1-z)^2 = 0$$

en donde z recorre la región que está a una distancia 1/8 del origen.

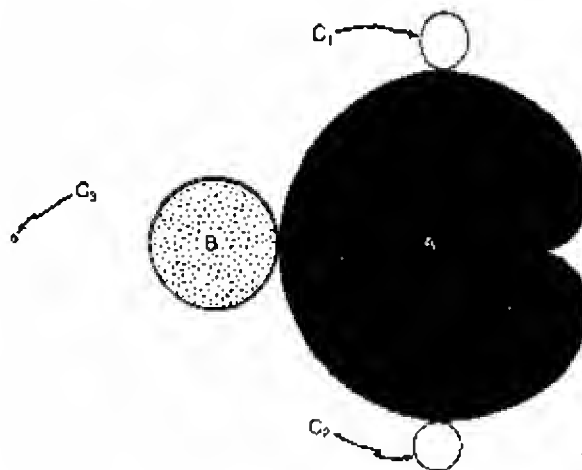


FIGURA IV. 13. *Las partes principales del interior del conjunto de Mandelbrot pueden ser definidas mediante sencillas ecuaciones algorítmicas.*

En realidad, esta ecuación nos da no solamente estas dos gotas (a la vez) sino también la forma "bebé" de tipo cardioide que aparece a la izquierda en la fig. III.2 —la región principal de la fig. III.1— y es la región marcada C_3 en la fig. IV. 13. de nuevo, estas regiones (juntas o separadas) constituyen conjuntos recursivamente enumerables (en el sentido sugerido antes), debido a la existencia de la fórmula anterior.

A pesar de la sugerencia en el sentido de que el conjunto de Mandelbrot puede ser no recursivo, ya hemos sido capaces de vaciar las áreas mayores del conjunto con algoritmos definidos y no-demasiado-complicados. Parece que este proceso continuará. Las regiones más visibles del conjunto —y ciertamente un porcentaje abrumador de su área (si no toda ella)— pueden tratarse

algorítmicamente. Si, como estoy suponiendo, el conjunto completo es realmente no recursivo, entonces las regiones que no pueden alcanzarse mediante nuestros algoritmos deben ser muy delicadas y difíciles de encontrar.

Más aún, cuando hayamos localizado una de estas regiones podremos mejorar nuestros algoritmos y llegar incluso a las regiones concretas. Pero entonces habría (si es correcta mi suposición de no recursividad) *otras* regiones ocultas aún más profundamente, tanto que ni siquiera nuestro algoritmo mejorado sería capaz de llegar. Una vez más, mediante esfuerzos prodigiosos de intuición, ingenio y aplicación, podríamos localizar una de estas regiones; pero todavía habría otras que se nos escaparían. Y así.

Nada de esto es distinto de la manera frecuente en que proceden las matemáticas sobre áreas en las que los problemas son difíciles y presumiblemente no recursivos. Los problemas más comunes pueden ser tratados con sencillos procedimientos algorítmicos, algunos de ellos conocidos durante siglos, pero otros escaparán de la red o requerirán de procedimientos más sofisticados. Por supuesto, los que aún escapen intrigarán particularmente a los matemáticos y les aguijonearán para desarrollar métodos más potentes, basados sobre cada vez más profundas reflexiones en torno a la naturaleza de las matemáticas implicadas. Hay algo de intuición en nuestra comprensión del mundo. En los problemas de la teselación y de las palabras subyace mucho de todo esto, no importa que se trate de áreas donde la maquinaria matemática no está aún muy desarrollada. Pudimos utilizar un argumento muy sencillo en cada caso particular, a fin de probar que una determinada palabra no puede ser obtenida a partir de otra mediante las reglas permitidas, pero no es difícil imaginar que entrarían en juego líneas arguméntales más sofisticadas para los casos complejos. Es muy probable que las nuevas líneas de razonamiento se desarrollen sobre la base de un procedimiento algorítmico.

Sabemos que ningún procedimiento es válido para todos los ejemplo[5]

del problema de las palabras, aunque los ejemplos que escapen tendrían que ser construidos con mucho cuidado y sutileza. En la práctica, cuando *sepamos* cómo se construyen esos ejemplos — cuando sepamos con certeza de un caso particular en que ha sido eludido nuestro algoritmo—, podremos mejorarlo e incluir también ese caso. Solamente pueden escapar pares de palabras que no son "iguales", de modo que en cuanto sepamos que han escapado, sabremos que no son "iguales". Y tal hecho puede ser añadido a nuestro algoritmo. Nuestra reflexión mejorada nos conducirá a un algoritmo mejorado.

TEORÍA DE LA COMPLEJIDAD

Los argumentos que he dado más arriba, y en los capítulos precedentes, respecto a la naturaleza, existencia y limitaciones de los algoritmos han sido en gran medida a un nivel "de principios". No he discutido en absoluto la cuestión de si los algoritmos que aparecen pueden ponerse en práctica. Incluso para problemas en donde sabemos que los algoritmos pueden construirse y conocemos cómo hacerlo, es necesario un trabajo arduo para desarrollarlos y aprovecharlos. A veces un poco de intuición e ingenio conduce a una reducción considerable en la complejidad, y a veces también a mejoras en su velocidad.

Estas cuestiones son con frecuencia demasiado técnicas y detalladas, y en los últimos años se ha trabajado en contextos diferentes para construir, comprender y mejorar los algoritmos —un área

de trabajo en rápida expansión y desarrollo—. No entraré en una discusión a detalle de esto pero se han establecido diversas generalidades, a veces no con base en meras conjeturas, acerca de ciertas limitaciones *absolutas* al incremento en la velocidad de un algoritmo. Inclusive se sabe que entre problemas matemáticos que *son* de naturaleza algorítmica existen algunas clases de problemas que son intrínsecamente más difíciles de resolver y que sólo pueden ser resueltos mediante algoritmos muy lentos (o, quizá, mediante algoritmos que requieren una cantidad desorbitadamente grande de espacio, etc.). La teoría que trata asuntos como éste se llama *teoría de la complejidad.*

La teoría de la complejidad se interesa no tanto en la solución algorítmica de problemas individuales como en familias infinitas de problemas de tal índole que habría un algoritmo general para todos los problemas de una sola familia. Los problemas de una familia son de diferentes "tamaños", y éstos son medidos por algún número natural n. (En seguida diré de qué manera.) La longitud de tiempo —o, más correctamente, el número de pasos elementales— que necesitará el algoritmo para cada problema particular vendrá dada por un número natural N que dependa de n. Esto quiere decir que, considerados *todos* los problemas de un tamaño particular n, el mayor número de pasos que necesitará el algoritmo será N. Ahora bien, a medida que n se hace más y más grande, es probable que el número N también se haga más y más grande, y hasta sería posible que N creciera mucho más rápidamente que aquél. Por ejemplo, N podría ser aproximadamente proporcional a n^2 o a n^3, o quizá a 2^n (que, para n grande, es mucho mayor que cada uno de los n, n^2, n^3, n^4 *y* n^5, mayor de hecho, que n^r *para* cualquier número r dado), e incluso N podría ser aproximadamente proporcional a, pongamos por caso, 2^{2^n} (que es todavía mucho mayor).

El número de "pasos" dependerá del tipo de máquina computadora en que sea ejecutado el algoritmo. Si la máquina computadora es una máquina de Turing del tipo descrito en el capítulo II, en el que sólo hay una cinta (lo que es ineficiente), entonces el número N podría crecer más rápidamente (es decir, la máquina podría funcionar más lentamente) que si se permitieran dos o más cintas. Para evitar equívocos como éste, se hace una amplia clasificación de las posibles maneras en las que crece N como función de n, de modo que —independientemente del tipo de máquina de Turing que se utilice— la medida de la tasa de crecimiento de N caiga siempre en la misma categoría. Una categoría semejante, denominada P (que significa "tiempo polinómico"), incluye todas las tasas, que son, como mucho, múltiplos fijos* de uno de los n, n^2, n^3, n^4, n^5,... O sea que para cualquier problema de la categoría P (donde por "problema" entiendo realmente una familia de problemas con un algoritmo general para resolverlos), tenemos

$$N < K \times n^r,$$

siendo los números K *y* r *constantes* (independientes de n), lo que significa que N no es mayor que un múltiplo de n elevado a una potencia fija.

Otro tipo de problemas que pertenecen a **P** es aquel que consiste en multiplicar dos números. Podemos imaginar que cada número se escribe en la notación binaria y que $n/2$ es el número de dígitos binarios de cada número, lo que dará un *total* de n dígitos binarios, es decir, n *bits*.(Si uno

* Un "polinomio" es una expresión más general, tal como $7n^4 - 3n^3 + 6n + 15$. Pero, en cualquiera de estas expresiones, los términos que incluyen potencias menores que n pierden importancia cuando n se hace grande (así que en este ejemplo concreto podemos ignorar todos los términos, excepto $7n^4$).

de los números es mayor que el otro, podemos hacer empezar el más corto con una sucesión de ceros para adaptarlo a la longitud del más grande.) Por ejemplo, si $n = 14$, consideraríamos

$$1011010 \times 0011011$$

(que es 1011010 x 11011, pero con ceros añadidos a la cifra más corta). El modo más directo de llevar a cabo esta multiplicación se escribiría:

```
        1011010
      x 0011011
      ---------
        1011010
       1011010
      0000000
     1011010
    1011010
   0000000
  0000000      .
  -------------
  0100101111110
```

y hay que recordar que, en el sistema binario, 0x0=0, 0x1=0, 1x0=0, 1x1=1, 0+0=0, 0+1=1, 1+0=1, 1+1=10.

El número de multiplicaciones binarias individuales es $(n/2) \times (n/2) = n^2/4$, y puede haber hasta $(n^2/4) - (n/2)$ sumas binarias individuales (inclusive las que se llevan de una columna a la siguiente). Esto hace $(n^2/2) - (n/2)$ operaciones aritméticas individuales en total, y añadiremos algunas extra para los pasos lógicos que implica el llevar una cifra en la suma. El número total de pasos es esencialmente $N = n^2/2$ (ignorando los términos de menor orden) que, ciertamente, es polinómico.[13]

En general, para una clase de problemas tomamos la medida n del "tamaño" del problema como el *número total de dígitos binarios* (o *bits*) que se requieren para especificar los datos libres del problema de ese tamaño particular. Esto significa que, para n dado, habrá hasta 2^n diferentes casos del problema (puesto que cada dígito puede ser una de las dos posibilidades: 0 o 1, y hay n dígitos en total), y éstos tienen que ser cubiertos uniformemente por el algoritmo, en no más de N pasos.

Existen muchos ejemplos de (clases de) problemas que *no* están en **P**. Por ejemplo, para realizar la operación del cálculo de 2^{2^r} a partir del número r necesitaremos alrededor de 2^n pasos tan sólo para *escribir la respuesta*, y ya no digamos para realizar el cálculo, siendo n el número de dígitos binarios en la representación binaria r. La operación de calcular $2^{2^{2^r}}$ necesita alrededor de 2^{2^r} pasos nada más para escribirla, etc. Son mucho mayores que los polinomios y, por consiguiente, no pertenecen a **P**.

Más interesantes son los problemas en los que las respuestas pueden escribirse, y aun verificarse dentro de un tiempo polinómico. Hay una categoría importante de problemas algorítmicamente resolubles que se caracterizan por esta propiedad. Se les llama (clases de) problemas **NP.** Si un problema individual de una clase de problemas en NP tiene solución, el algoritmo dará esa solución, y además será posible verificar en un tiempo polinómico que la solución propuesta es

[13] De hecho, mediante el uso del ingenio, este número de pasos puede reducirse a algo orden de $n \log n \log \log n$ para n grande que, por supuesto, sigue estando en P. Véase Knuth (1981) para más información sobre estas materias.

realmente eso: una solución. En los casos en que el problema no tiene solución, el algoritmo lo dirá, pero no se exige que verifique —en tiempo polinómico o en cualquier otro— que no hay tal solución.[14]

Los problemas NP aparecen no sólo en el contexto de las matemáticas, sino en muchos otros. Sólo daré un ejemplo matemático: el problema de encontrar lo que se denomina un *circuito hamiltoniano* (un nombre rimbombante para una idea extremadamente simple) en un grafo, entendiendo éste como una colección finita de puntos, o vértices, de los cuales un cierto número de pares está conectado mediante líneas —llamadas las "aristas" del grafo—. (No estamos interesados aquí en propiedades geométricas o de "distancia", sino sólo en los vértices que están conectados entre sí. Por consiguiente, no importa que los vértices estén representados en un plano —ni que las aristas se crucen— o en un espacio tridimensional.) Un circuito hamiltoniano es un lazo o un camino cerrado que sólo consta de las aristas del grafo y que pasa solamente una vez por cada vértice (véase la fig. IV. 14). El problema estriba en determinar, para cualquier grafo dado, si existe o no un circuito hamiltoniano, y en representarlo explícitamente en el caso de que exista. Hay varias maneras de presentar un grafo en términos de dígitos binarios. No importa mucho cuál sea el método utilizado. Un procedimiento sería numerar los vértices 1, 2, 3, 4, 5,... y hacer luego una lista de los pares en un orden fijo adecuado. Por ejemplo:

(1, 2), (1, 3), (2, 3), (1, 4), (2, 4), (3, 4), (1, 5), (2, 5), (3, 5), (4, 5), (1, 6),...

A continuación hacemos una lista en la que ponemos un "1" allí donde el par corresponde a una arista y un "0" si no lo hace. Así, la secuencia binaria

10010110110...

significará que el vértice 1 está unido al vértice 2, al vértice 4 y al vértice 5,... el vértice 3 está unido al vértice 4 y al vértice 5,... el vértice 4 está unido al vértice 5,... etc. (como en la fig. IV. 14).

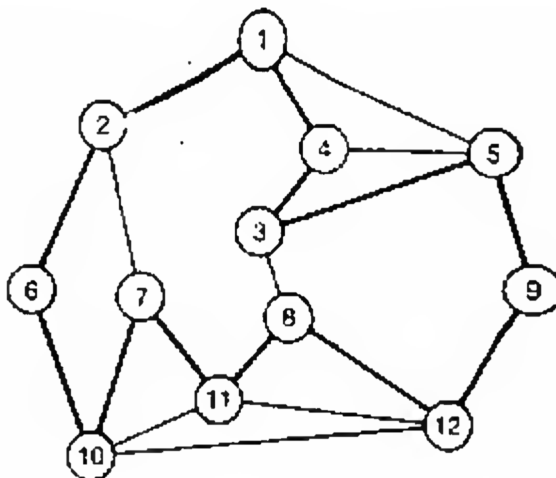


FIGURA IV. 14. *Un grafo con un circuito hamiltoniano indicado: líneas más oscuras. Existe un circuito hamiltoniano más que el lector puede tratar de localizar.*

[14] Más precisamente, las clases P, NP y NP completo son definidas para problemas de tipo sí/no (*v. gr.* dados *a, b y c*, ¿se verifica *a* x *b* = *c*?), pero las descripciones mostradas en el texto son adecuadas para nuestros propósitos.

El circuito hamiltoniano podría venir dado, si quisiéramos, simplemente como una subcolección de aristas descrita mediante una secuencia binaria con muchos más ceros que antes. El procedimiento de verificación es algo que puede conseguirse mucho más rápidamente que el descubrimiento inicial del circuito hamiltoniano. Todo lo que necesitamos hacer es verificar que el circuito propuesto es realmente un circuito, que sus aristas pertenecen a las del grafo original y que cada vértice del grafo se utiliza exactamente dos veces, cada una de ellas en los extremos de dos aristas distintas. Este procedimiento de verificación puede hacerse en un tiempo polinómico.

Este problema no sólo es **NP,** sino que se conoce como **NP** completo, y ello significa que cualquier otro problema **NP** puede ser reducido a este en un tiempo polinómico —de modo que si alguien fuera lo bastante hábil para encontrar un algoritmo que resolviese el problema del circuito hamiltoniano en tiempo *polinómico*, es decir, que probara que el problema del circuito hamiltoniano está en P, se concluiría que *todos* los problemas **NP** están realmente en **P**. Tal circunstancia tendría extraordinarias implicaciones. De un modo general, los problemas que están en **P** se consideran "tratables" (es decir, hallan solución en un intervalo aceptable), para n razonablemente grandes, en una rápida computadora moderna, mientras que los problemas en NP que no están en P se consideran "intratables" (es decir, aunque solubles en principio, sin solución en la práctica) para n razonablemente grande independientemente de los incrementos, de cualquier tipo previsible, en cualquier velocidad imaginaria de la computadora. (El tiempo real que se necesitaría para un problema **NP** no en **P**, para n grande, se haría velozmente mayor que la edad del universo, lo que no sirve de mucho en un problema práctico). Y cualquier algoritmo para solucionar el problema del circuito hamiltoniano en un tiempo polinómico podría transformarse en un algoritmo para resolver *cualquier* otro problema **NP** en un tiempo igual.

Otro problema que es **NP** completo[15] es el "problema del viajante", muy similar al problema del circuito hamiltoniano excepto porque las diversas aristas tienen números asociados a ellas y se busca el circuito hamiltoniano para el que la suma de los números (la "distancia" recorrida por el viajante) es *mínima*. Una vez más, una solución en un tiempo polinómico conduciría a una solución en tiempo igual para todos los demás problemas **NP.** (Encontrar tal solución constituiría una noticia de portada. Existen sistemas secretos de codificación, introducidos en los últimos años, que dependen de un problema de factorización de números enteros altos, siendo éste otro problema **NP.** Si este problema pudiera solucionarse en tiempo polinómico, entonces tales códigos podrían ser probablemente descifrados con ayuda de los potentes ordenadores modernos, pero si no es así, los códigos estarán a salvo. Véase Gardner, 1989.)

Es opinión común de los expertos que es *imposible* resolver, con cualquier dispositivo similar a una máquina de Turing, un problema **NP** completo en tiempo polinómico y que, por consiguiente, **P** y **NP** *no* son el mismo. Es probable que tal creencia sea correcta, pero todavía nadie ha sido capaz de demostrarlo. Este sigue siendo el más importante problema no resuelto de la teoría de la complejidad.

[15] Estrictamente, necesitamos una versión sí/no de este problema, como "¿hay ruta para el viajante más corta que ésta o aquélla? (Véase *supra* nota 14.)

COMPLEJIDAD Y COMPUTABILIDAD EN LOS OBJETOS FÍSICOS

La teoría de la complejidad es importante para nuestras consideraciones en este libro, debido a que plantea un asunto un tanto desligado de la cuestión de si las cosas son o no son algorítmicas, a saber: el de si las cosas que sabemos que son algorítmicas lo son o no de un modo *útil*.

En los últimos capítulos tendré menos que decir acerca de la teoría de la complejidad que de la computabilidad. Ahora me inclino a pensar que a diferencia de la misma cuestión básica de la computabilidad, los resultados de la teoría de la complejidad no son cruciales en relación con los fenómenos mentales. Más aún, tengo la sensación de que las cuestiones de la factibilidad de los algoritmos apenas tienen algo en común con la teoría de la complejidad tal como hoy existe.

Sin embargo, pudiera estar equivocado con relación al papel de la complejidad. Como señalaré más adelante (en el capítulo IX), la teoría de la complejidad para *objetos físicos reales* podría diferir en aspectos significativos de la que hemos estado discutiendo, y para que esa posible diferencia se haga manifiesta es necesario aprovechar las propiedades "mágicas" de la mecánica cuántica —una misteriosa pero poderosa y precisa teoría del comportamiento de átomos y moléculas, y de muchos otros fenómenos, algunos de los cuales son importantes a una escala mayor. Aprenderemos algo de esta teoría en el capítulo VI.

Según ideas recientes de David Deutsch (1985), es posible *en principio* construir una computadora cuántica para la que existen (clases de) problemas que no están en P, pero que podrían ser resueltos por dicho dispositivo en tiempo polinómico. No está claro todavía cómo podría construirse un dispositivo físico confiable que se comporte (confiablemente) como una computadora cuántica —y, además, la clase particular de problemas considerada hasta ahora es decididamente artificial—, pero subsiste la posibilidad *teórica* de que un dispositivo físico cuántico mejoraría una máquina de Turing.

¿Sería posible que un cerebro humano —que para nuestro estudio estoy considerando como un "dispositivo físico" sorprendentemente sutil, delicado en su diseño, así como complicado— estuviera sacando provecho de la teoría cuántica? ¿Comprendemos el modo en que podrían ser aprovechados los efectos cuánticos para la solución de problemas y la formación de juicios? ¿Es concebible que tengamos que ir aún "más allá" de la teoría cuántica de hoy para hacer uso de esas ventajas? ¿En verdad los dispositivos físicos pueden mejorar la teoría de la complejidad para máquinas de Turing? ¿Qué sucede con la teoría de la *computabilidad* para dispositivos físicos reales?

Para abordar estos temas debemos apartarnos de lo puramente matemático y preguntar, en los próximos capítulos: ¿cómo se comporta realmente el mundo físico?

V. EL MUNDO CLÁSICO

EL STATUS DE LA TEORÍA FÍSICA

¿QUE HAY QUE CONOCER del funcionamiento de la naturaleza para poder apreciar cómo la conciencia puede formar parte de ella? ¿De veras importa cuáles son las leyes que gobiernan los elementos que constituyen el cuerpo y el cerebro? Si nuestras percepciones conscientes consistieran simplemente en la activación de algoritmos, como pretenden muchos defensores de la IA, entonces no importaría gran cosa cuáles son esas leyes. Cualquier dispositivo que sea capaz de ejecutar un algoritmo sería tan bueno como cualquier otro. Quizá, por el contrario, haya más que simples algoritmos en nuestras sensaciones de conciencia. Tal vez sea importante nuestra muy particular constitución, así como las leyes físicas que gobiernan la sustancia de que estamos compuestos. Quizá necesitemos comprender cuál es la cualidad profunda en la naturaleza misma de la materia y qué determina la manera como esta materia debe comportarse. La física no ha llegado todavía a este punto. Aún quedan muchos misterios que desentrañar y muchas intuiciones que obtener. Sin embargo, muchos físicos y fisiólogos dirían que *ya* sabemos bastante sobre las leyes físicas que rigen el funcionamiento de un objeto como el cerebro humano. Aunque es indudable que éste es excepcionalmente complicado como sistema físico, y que aún no conocemos mucho de su estructura y comportamiento, pocos estarían dispuestos a afirmar que es precisamente en los principios *físicos* que determinan su comportamiento donde existe una considerable falta de comprensión.

Más adelante defenderé el planteamiento poco convencional de que, contrariamente a esta opinión, todavía *no* comprendemos suficientemente bien la física como para describir adecuadamente en sus términos el funcionamiento de nuestro cerebro, ni siquiera en principio. Para hacer este planteamiento, será necesario presentar en primer lugar una visión general de la teoría física actual. Este capítulo concierne a la llamada "física clásica", que comprende tanto la mecánica de Newton como la relatividad de Einstein. Aquí, por "clásico" entendemos esencialmente las teorías dominantes antes de la llegada, alrededor de 1925 (mediante el trabajo inspirado de físicos como Planck, Einstein, Bohr, Heisenberg, Schrödinger, De Broglie, Born, Jordan, Pauli y Dirac), de la *teoría cuántica*, con su incertidumbre, indeterminismo y misterio, teoría describe el comportamiento de las moléculas, los átomos y las partículas subatómicas. La teoría clásica es, por el contrario, *determinista* por cuanto supone que el futuro siempre está completamente condicionado por el pasado. Aun así, la física clásica tiene mucho de misterioso, pese al hecho de que el conocimiento a que ha dado lugar a lo largo de siglos nos ha llevado a una imagen de gran precisión fenoménica. Tendremos que examinar también la teoría cuántica (en el capítulo VI) pues, contrariamente a lo que parece ser la opinión mayoritaria entre los fisiólogos, *es* probable que los fenómenos cuánticos sean importantes en el modo de operar del cerebro, pero esto es tema para los siguientes capítulos.

Lo que la ciencia ha conseguido hasta ahora ha sido espectacular. Sólo tenemos que mirar a nuestro alrededor para atestiguar lo que el extraordinario poder de nuestra comprensión de la naturaleza nos ha ayudado a obtener. La tecnología del mundo moderno se ha derivado, en buena medida, de una gran riqueza de experiencias empíricas. Sin embargo, es la *teoría* física la que fundamentalmente sustenta nuestra tecnología, de manera que es dicha teoría la que aquí nos interesará. La precisión de las teorías de las que disponemos ahora es bastante notable. Pero no es sólo su precisión la que les da fuerza. También lo es el hecho de que se han mostrado extraordinariamente susceptibles de un tratamiento matemático preciso y detallado. Estos hechos en conjunto han proporcionado una ciencia verdaderamente impresionante por su vigor.

Parte considerable de esta teoría física no es tan reciente. Si ha de señalarse un suceso sobre todos los demás, éste es la publicación, en 1687, de los *Principia* de Isaac Newton. Esta obra extraordinaria demostró cómo, a partir de unos pocos principios físicos, se puede comprender, y a menudo pronosticar con sorprendente precisión, gran parte del comportamiento real de los objetos físicos. (Una buena porción de los *Principia* tenía que ver también con notables avances en los métodos matemáticos, aunque más tarde Euler y otros proporcionaron métodos más prácticos.) El propio trabajo de Newton, como él fue el primero en admitir, debía mucho a los pensadores anteriores, entre los que destacaban los nombres de Galileo Galilei, Rene Descartes y Johannes Kepler. Pero había importantes conceptos que procedían de pensadores todavía más antiguos, como las ideas geométricas de Platón, Eudoxo, Euclides, Arquímedes y Apolonio. Más adelante habrá más que decir al respecto.

Las desviaciones del esquema básico de la dinámica de Newton vendrían más tarde. La primera fue la teoría electromagnética de James Clerk Maxwell, desarrollada a mediados del siglo XIX, y que englobaba no sólo el comportamiento clásico de los campos eléctrico y magnético sino también el de la luz.[1] Esta importante teoría será objeto de nuestra atención más adelante, en este capítulo. La teoría de Maxwell es de considerable importancia para la tecnología actual, y no hay duda de que los fenómenos electromagnéticos tienen que ver con el funcionamiento de nuestro cerebro. Lo que es menos evidente, sin embargo, es que también pueden ser importantes para nuestros procesos mentales las dos grandes teorías de la relatividad asociadas con el nombre de Albert Einstein. La teoría *especial* de la relatividad, que surgió a partir de un estudio de las ecuaciones de Maxwell, fue desarrollada por Herny Poincaré, Hendrick Antoon Lorentz y Einstein (y más tarde Hermann Minkowski hizo de ella una elegante descripción geométrica) para explicar el enigmático comportamiento de los cuerpos cuando se mueven a velocidades próximas a la de la luz. La famosa ecuación de Einstein $E = mc^2$ era parte de esta teoría. Pero el efecto de la teoría en la tecnología ha sido muy pequeño (excepto allí donde incide sobre la física nuclear) y su importancia para el funcionamiento de nuestro cerebro parecería ser, en el mejor de los casos, marginal. No obstante, la relatividad especial nos dice algo profundo sobre la realidad física, en relación con la naturaleza del *tiempo*. Veremos en los próximos capítulos que esto conduce a profundos enigmas de la teoría cuántica que podrían ser de importancia para nuestra percepción del "flujo del tiempo". Además, necesitaremos comprender la teoría especial para poder estimar debidamente la teoría *general* de la relatividad de Einstein —la cual utiliza el espacio-tiempo curvo para describir la gravedad—. Hasta ahora *esta* teoría casi no ha influido en la tecnología[*] y parecería en extremo fantasioso pretender que tuviera algo que ver con el funcionamiento de nuestro cerebro. Pero, curiosamente, es en realidad la teoría *general* la que será tomada en cuenta nuestros planteamientos posteriores, particularmente en los capítulos VII

[1] Es sorprendente el hecho de que todas las desviaciones establecidas de la imagen newtoniana han estado esencialmente asociadas con el comportamiento de la luz. En primer lugar, están los campos incorpóreos que transportan energía en la teoría electromagnética de Maxwell. En segundo lugar está, como veremos, el papel primordial que desempeña la velocidad de la luz en la teoría de la relatividad especial de Einstein. En tercer lugar, las pequeñísimas desviaciones de la teoría gravitatoria de Newton que exhibe la teoría de la relatividad general de Einstein, se vuelven significativas sólo cuando las velocidades son comparables con la de la luz. (La desviación de la luz por el Sol, el movimiento de Mercurio, las velocidades de escape comparables a la de la luz para agujeros negros, etc.) En cuarto lugar está la dualidad onda-corpúsculo de la teoría cuántica, observada en primer lugar en el comportamiento de la luz. Finalmente está la electrodinámica cuántica, que es la teoría cuántica de la luz y las partículas cargadas. Es razonable suponer que el propio Newton habría estado dispuesto a aceptar que su imagen del mundo enfrentaría profundos problemas ocultos tras el misterioso comportamiento de la luz (cfr. Newton, 1730; también Penrose, 1987a).

[*] Sin embargo, la precisión exigida para el comportamiento de las sondas espaciales requiere que sus órbitas se calculen tomando en consideración los efectos de la relatividad general; y existen dispositivos capaces de localizar una posición en la Tierra con tal exactitud (de hecho, con un error de apenas unos decímetros) que deben tener en cuenta los efectos de la curvatura del espacio-tiempo de la relatividad general.

y VIII, en los que tendremos que aventurarnos en los más remotos confines del espacio y el tiempo para hacernos una idea de los cambios que considero necesarios antes de que pueda salir a la luz una imagen coherente de la teoría cuántica.

Estas son las áreas generales de la física *clásica*. ¿Qué pasa con la física cuántica? A diferencia de la teoría de la relatividad, la teoría cuántica está empezando a tener repercusiones importantes en la tecnología. Esto se debe en parte a los conocimientos que ha proporcionado en ciertas áreas tecnológicamente importantes, como la química y la metalurgia. Algunos dirán que estas áreas han quedado incluidas dentro de la física en virtud de las nuevas intuiciones que nos ha brindado la teoría cuántica. Además de esto, la teoría cuántica nos ha llevado a muchos fenómenos *nuevos*, el más familiar de los cuales es, supongo, el láser. ¿No podría ser que algunos aspectos esenciales de la teoría cuántica desempeñaran también un papel determinante en la física que rige nuestros procesos mentales?

¿Qué hay de los conocimientos físicos más recientes? Algunos lectores se habrán topado con emocionantes conceptos como el de los *quarks*, las GUT (Teorías de la Gran Unificación), el "escenario inflacionario" (véase la nota 13 del capítulo VII), la "supersimetría", la "teoría de las (super) cuerdas", etc. ¿Cómo se comparan estos nuevos esquemas con los que acabo de mencionar? ¿Necesitaremos saber también algo sobre ellos? Para colocar las cosas en una perspectiva más apropiada, dividiré en tres amplias categorías las teorías físicas básicas. A saber:

1. SUPREMAS
2. ÚTILES
3. PROVISIONALES

En la categoría de SUPREMAS deben ir todas las que hemos considerado en los párrafos anteriores. Para calificarlas de SUPREMAS no estimo necesario que la teoría se aplique sin refutación a los fenómenos del mundo; sólo exijo que el alcance y exactitud con que se apliquen sea *excepcional*, en el sentido apropiado. Tal como empleo el término "suprema", resulta extraordinariamente notable el simple hecho de que existan teorías dentro de esta categoría. No conozco ninguna teoría básica de ninguna otra ciencia que pudiera encajar debidamente en esta categoría. Quizá la teoría de la selección natural, que propusieron Darwin y Wallace, sea la que está más próxima aunque todavía a una buena distancia. La más antigua de las teorías SUPREMAS es la geometría euclidiana de la que algo aprendemos en la escuela. Las antiguos pudieron no considerarla como una teoría física en absoluto, pero eso es lo que realmente es: una teoría sublime y supremamente precisa del espacio físico (y de la geometría de los cuerpos rígidos). ¿Por qué considero la geometría euclidiana una teoría *física* en lugar de una rama de las matemáticas? Irónicamente, una de las razones más evidentes para adoptar esta opinión es que ahora sabemos que la geometría euclidiana *no es completamente exacta* como descripción del espacio físico en que habitamos. La teoría de la relatividad general de Einstein nos dice ahora que el espacio(-tiempo) es realmente "curvo" (es decir, *no* exactamente euclidiano) cuando se enmarca en un campo gravitatorio. Pero este hecho no invalida el calificar de SUPREMA a la geometría euclidiana. Para dimensiones en la escala del metro, las desviaciones respecto a la planitud euclidiana son ínfimas, y los errores al tratar la geometría como euclidiana son menores que el diámetro de un átomo de hidrógeno.

Es razonable decir que también debería calificarse de SUPREMA la teoría de la *estática* (que trata de los cuerpos en reposo), tal como la desarrollaron Arquímedes, Pappus y Stevin en forma de una hermosa ciencia. Esta teoría se incluye ahora en la mecánica newtoniana. Las ideas profundas de la *dinámica* (es decir, de los cuerpos en movimiento) introducidas por Galileo alrededor del 1600, desarrolladas por Newton hasta constituir una magnífica y amplia teoría, deben entrar indudablemente en la categoría de SUPREMAS. Cuando se aplica a los movimientos de los planetas y satélites, la precisión observada de la teoría es excepcional: con un margen de error inferior a una parte en diez millones. El mismo esquema newtoniano se aplica aquí en la Tierra —o entre las estrellas y galaxias— con una precisión comparable. Análogamente, la teoría de Maxwell es válida con gran exactitud dentro de un vastísimo dominio que en el extremo inferior se extiende hasta la minúscula escala de los átomos y partículas subatómicas, y en el superior, hasta la escala de las galaxias —aproximadamente un millón de millones de millones de millones de millones de millones de veces mayor. (En el extremo más pequeño de esta escala las ecuaciones de Maxwell deben combinarse adecuadamente con las reglas de la mecánica cuántica.) Ciertamente, esta teoría también debe ser calificarse de SUPREMA.

La relatividad especial de Einstein (anticipada por Poincaré y elegantemente reformulada por Minkowski) da una descripción maravillosamente precisa de los fenómenos en los que la velocidad de los objetos llega a ser próxima a la de la luz —velocidades a las que las descripciones de Newton comienzan a fallar—. La teoría de la relatividad general de Einstein, de suprema belleza y originalidad, generaliza la teoría dinámica de Newton y su concepto de gravedad, y mejora su exactitud, a la vez que hereda, toda la notable precisión de esta teoría respecto al movimiento de los planetas y satélites. Además, explica con detalle diversos hechos observacionales que son incompatibles con el esquema newtoniano. Uno de estos hechos (el "pulsar binario",) muestra que la teoría de Einstein tiene una exactitud de alrededor de una parte en 10^{14}. Ambas teorías de la relatividad —la segunda de las cuales incluye a la primera— deben clasificarse realmente como SUPREMAS (casi tanto en razón de su elegancia matemática como en virtud de su exactitud).

La diversidad de fenómenos que se explican conforme a la extrañamente bella y revolucionaria teoría de la mecánica cuántica, la exactitud con que concuerda con los experimentos, nos dice claramente que también la teoría cuántica debe ser calificada de SUPREMA. No se conocen discrepancias observacionales con dicha teoría, aunque su fuerza reside, más allá de esto, en el número de fenómenos antes inexplicables y que la teoría explica ahora. Las leyes de la química, la estabilidad de los átomos, la agudeza de las líneas espectrales y sus muy específicas estructuras observadas, el curioso fenómeno de la superconductividad (resistencia eléctrica nula) y el comportamiento de los láseres son sólo algunos de éstos.

Son altos los requisitos que establecemos para que una teoría ingrese en la categoría de SUPREMA, pero en la física nos hemos acostumbrado a esto. Ahora bien, ¿qué hay sobre las teorías más recientes? En mi opinión sólo hay una que pueda calificarse de SUPREMA, y no es particularmente reciente: la teoría llamada *electrodinámica cuántica* (o EDC), que surgió del trabajo de Jordán, Heisenberg y Pauli, fue formulada por Dirac en 1926-1934, y hecha manejable por Bethe, Feynman, Schwinger y Tomonaga en 1947-1948. Esta teoría apareció como una combinación de los principios de la mecánica cuántica con los de la relatividad especial, incorporando las ecuaciones de Maxwell y una ecuación fundamental, debida a Dirac, que gobierna el movimiento y el *spin* del electrón. La teoría en conjunto no tiene la irresistible elegancia ni la solidez de las teorías SUPREMAS anteriores, pero se califica así en virtud de su

precisión verdaderamente excepcional. Un resultado particularmente digno de mención es el valor del momento magnético del electrón. (Los electrones se comportan como minúsculos imanes de carga eléctrica en rotación. La expresión "momento magnético" se refiere a la fuerza de este minúsculo imán.) El valor calculado para este momento magnético a partir de la EDC es 1.001 159 652 46 (en las unidades apropiadas, con un margen de error de alrededor de 20 en las dos últimas cifras), mientras que el valor experimental más reciente es 1.001 159 652 193 (con un posible error de alrededor de 10 en las dos últimas cifras). Como ha señalado Feynman, esta precisión equivale a determinar la distancia entre Nueva York y Los Ángeles ¡con un error menor que el espesor de un cabello humano! No tendremos aquí necesidad de conocer esta teoría pero, para dar una visión más cabal, mencionaré brevemente algunas de sus características fundamentales hacia el final del próximo capítulo.[*]

Existen algunas teorías actuales que yo colocaría en la categoría de ÚTILES. Aunque dos de éstas no serán aquí necesarias, sí son dignas de mención. La primera es la del modelo de *quarks* de Gell-Mann-Zweig para las partículas subatómicas llamadas *hadrones* (los protones, neutrones, mesones, etc., que constituyen los núcleos atómicos o, más correctamente, las partículas "fuertemente interactivas") y la (posterior) teoría detallada de sus interacciones, conocida como *cromodinámica cuántica* o *CDC*. La idea consiste en que todos los hadrones están formados por constituyentes conocidos como "quarks" que interactúan entre sí mediante una cierta generalización de la teoría de Maxwell (llamada teoría de Yang-Mills). En segundo lugar, existe una teoría (debida a Glashow, Salam, Ward y Weinberg, y que utiliza una vez más la teoría de Yang-Mills) que combina las fuerzas electromagnéticas con las interacciones "débiles" que son las responsables de la desintegración radioactiva. Esta teoría incorpora una descripción de los llamados *leptones* (electrones, muones, neutrinos; también de las partículas *W* y *Z*, las partículas "débilmente interactivas"). Hay un buen fundamento experimental para ambas teorías. Sin embargo, ellas son, por varias razones, algo más desordenadas de lo que uno quisiera (como sucedía con la EDC, pero más en este caso) y su exactitud observada y poder predictivo quedan, por el momento, a mucha distancia del nivel "excepcional" que se exige para incluirlas en la categoría de SUPREMAS. Estas dos teorías juntas (la segunda incluye la EDC) se conocen a veces como el *modelo estándar*.

Finalmente, existe una teoría de otro tipo que creo también pertenece, cuando menos, a la categoría de ÚTILES. Esta es la teoría llamada del *big bang* o *gran explosión*, sobre el origen del Universo.[**] Esta teoría desempeñará un papel importante en los capítulos VII y VIII.

No creo que ninguna otra teoría pueda entrar en la categoría de ÚTILES.[2] Existen muchas ideas hoy populares. Algunas de ellas son: las teorías de Kaluza-Klein, las de la "supersimetría" (o

---

[*] Véase el libro *QED* de Feynman (1985), donde se ofrece una exposición simplificada de esta teoría.

[**] Me refiero aquí a lo que se conoce como el "modelo estándar" de la gran explosión. Existe muchas variantes de esta teoría las más populares proporcionan actualmente lo que se conoce como "escenario inflacionario". En mi opinión, están claramente en la categoría de PROVISIONALES.

[2] Hay un magnífico cuerpo de conocimientos físicos bien establecido —la *termodinámica* de Carnot, Maxwell, Kelvin, Boltzmann y otros— que he dejado sin clasificar. Aunque esto puede intrigar a algunos lectores, la omisión ha sido deliberada. Por razones que se harán evidentes en el capítulo VII, yo mismo sería bastante reacio a colocar la termodinámica, tal como está, en la categoría de las teorías SUPREMAS. Sin embargo, muchos físicos considerarán probablemente un *sacrilegio* colocar un cuerpo de ideas tan bello y fundamental en una categoría tan modesta como la de simplemente ÚTILES. En mi opinión, la termodinámica, como se entiende normalmente, siendo algo que se aplica solamente a *promedios*, y no a los constituyentes individuales de un sistema —y siendo parcialmente una deducción de otras teorías— no es una teoría física en el sentido en que lo entiendo aquí (lo mismo se aplica a la estructura matemática de la mecánica estadística). Uso este hecho como excusa para evitar el problema y dejarlo fuera de la clasificación. Como veremos en el capítulo VII, afirmo que existe una íntima conexión

"supergravedad"), y las teorías ahora muy de moda de las "cuerdas" (o "supercuerdas"), además de las teorías GUT (y ciertas ideas derivadas de ellas, como el "escenario inflacionario", *cfr.* nota 13). Todas ellas entran de lleno, a mi modo de ver, en la categoría de PROVISIONALES. (Véase Barrow, 1988; Close, 1983; Davies y Brown, 1988; Squires, 1985.) La diferencia importante entre las categorías de ÚTILES y PROVISIONALES es la falta de cualquier fundamento experimental importante para las teorías de esta última categoría.[3] Esto no quiere decir que alguna de ellas no pudiera ascender a la categoría de las ÚTILES o incluso a la de SUPREMAS. Algunas de estas teorías contienen ideas originales muy prometedoras, pero por ahora siguen siendo ideas sin fundamento experimental. La categoría de las PROVISIONALES es una categoría muy amplia. Las ideas implícitas en algunas de ellas podrían contener las semillas de un nuevo avance sustancial en el conocimiento, mientras que algunas otras me dan la impresión de ser artificiosas o estar de plano descaminadas. (Me vi tentado a formular una cuarta categoría a partir de la respetable categoría de PROVISIONALES y llamarla, por ejemplo, DESCAMINADAS; pero luego lo pensé mejor, pues no quiero perder a la mitad de mis amigos.)

No debería sorprendernos que las principales teorías SUPREMAS sean antiguas. A lo largo de la historia debe haber habido muchas más teorías que entrarían en la categoría de PROVISIONALES, pero la mayoría de ellas han sido olvidadas. Análogamente, debe haber habido otras muchas en la categoría de ÚTILES que se han desvanecido desde entonces; pero había también algunas que se han incorporado en teorías que más tarde llegaron a ser SUPREMAS por sí mismas. Consideremos unos pocos ejemplos. Antes de que Copérnico, Kepler y Newton concibieran un esquema mucho mejor, existía una teoría del movimiento planetario maravillosamente elaborada que habían desarrollado los antiguos griegos, conocida como *sistema tolemaico*. Según este esquema los movimientos de los planetas están gobernados por complicadas composiciones de movimientos circulares. Fue bastante eficaz para hacer predicciones, pero se hizo más y más complicado a medida que se necesitaba mayor exactitud. Hoy día el sistema tolemaico nos parece muy artificioso.

Este es un buen ejemplo de una teoría ÚTIL (lo fue de hecho durante unos veinte siglos) que posteriormente se *disolvió* como teoría física aunque tuvo un papel organizativo de clara importancia histórica. Como un buen ejemplo de teoría ÚTIL del tipo finalmente *acertado* podemos considerar, en su lugar, la brillante concepción de Kepler del movimiento planetario elíptico. Otro ejemplo fue la tabla periódica de Mendeleyev para los elementos químicos. Por sí mismas, no proporcionaban esquemas productivos con el carácter "excepcional" exigido, pero posteriormente llevaron a hacer deducciones "correctas" dentro de teorías SUPREMAS que se desarrollaron a partir de ellas (la dinámica newtoniana y la teoría cuántica, respectivamente).

En las secciones y capítulos siguientes no tendré mucho qué decir sobre las teorías actuales que son simplemente ÚTILES o PROVISIONALES. Hay bastante qué decir sobre las SUPREMAS. Es un hecho afortunado que tengamos tales teorías y podamos comprender el mundo en que vivimos de una forma tan completa. Con el tiempo debemos tratar de averiguar si incluso estas teorías son

---

entre la termodinámica y un tema que he citado antes dentro de la categoría de ÚTILES; a saber, el modelo estándar de la gran explosión. Según creo, deberíamos considerar una unión apropiada entre estos dos conjuntos de ideas (que en parte falta actualmente) como una teoría física en el sentido exigido —incluso perteneciente a la categoría de SUPREMAS—. Sobre esto habremos de volver más adelante.

[3] Mis colegas me han preguntado dónde colocaría la "teoría de los *twistors*" —elaborada colección de ideas y procedimientos con la que he estado relacionado durante muchos años—. En la medida en que una teoría de *twistor* es una teoría diferente sobre el mundo físico, no puede estar más que en la categoría de PROVISIONALES; pero en buena medida no es en absoluto una teoría, sino una descripción matemática de teorías físicas previamente bien establecidas.

suficientemente ricas para gobernar las acciones del cerebro y la mente humanos. Traeré a colación esta cuestión a su debido tiempo, pero por ahora consideraremos las teorías SUPREMAS tal como las conocemos e intentaremos ponderar su importancia para nuestros propósitos.

LA GEOMETRÍA EUCLIDIANA

La geometría euclidiana no es más que esa materia que aprendemos como "geometría" en la escuela. Sin embargo, supongo que la mayoría de la gente la considera parte de las matemáticas, más que como una teoría física. Por supuesto es también parte de las matemáticas, pero la geometría euclidiana no es ni con mucho la única geometría matemática concebible. La geometría que nos fue transmitida por Euclides describe con gran exactitud el espacio físico del mundo en que vivimos, pero *no* es una necesidad lógica; es sólo una característica (aproximadamente exacta) *observada* del mundo físico.

De hecho, hay otra geometría, llamada *lobachevskiana** (o *hiperbólica*) que es en muchos aspectos muy similar a la geometría euclidiana pero con algunas curiosas diferencias. Por ejemplo, recordamos que en la geometría euclidiana la suma de los ángulos de cualquier triángulo es siempre 180°. En la geometría lobachevskiana esta suma es siempre *menor* de 180°, siendo la diferencia proporcional al área del triángulo (véase fig. V.1).

El famoso artista holandés Maurits C. Escher ha concebido algunas representaciones muy hermosas y exactas de esta geometría. En la fig. V.2 se reproduce uno de sus grabados. Cada pez negro debe ser imaginado, según la geometría lobachevskiana, del mismo tamaño y forma que cualquiera de los otros peces negros, y lo mismo es válido para los peces blancos. La geometría no puede representarse de forma completamente exacta en el plano euclidiano ordinario; de ahí el aparente apiñamiento en las proximidades del contorno circular. Imagínese usted mismo situado en el interior de la figura pero en algún lugar próximo a este contorno; se supone entonces que el espacio lobachevskiano se ve igual desde el centro, que de cualquier otro lugar. Lo que parece ser el "contorno" de la estructura, según esta representación euclidiana, está realmente, para la geometría lobachevskiana, "en el infinito". El contorno circular no debe considerarse en absoluto como parte del espacio lobachevskiano —y tampoco como parte de la región euclidiana en el exterior de este círculo—. (Esta ingeniosa representación del plano de Lobachevsky se debe a Poincaré. Tiene la virtud especial de que las formas muy pequeñas no quedan distorsionadas en la representación; sólo los tamaños cambian.) Las "líneas rectas" de la geometría (a lo largo de algunas de las cuales apuntan los peces de Escher) son círculos que intersectan en ángulos rectos este contorno circular.

Podría muy bien suceder que la geometría lobachevskiana fuera realmente verdadera en nuestro mundo a escala cosmológica (véase capítulo VII). Sin embargo, la constante de proporcionalidad entre el déficit de ángulo para un triángulo y su área tendría que ser *extraordinariamente* pequeña en este caso, y la geometría euclidiana sería una excelente aproximación a esta geometría para cualquier escala ordinaria. De hecho, como veremos más adelante en este mismo capítulo, la teoría de la relatividad general de Einstein nos dice que la geometría de nuestro mundo *difiere* de la geometría euclidiana (aunque de un modo "irregular" que es más complicado

* Nicolai Ivanovich Lobachevsky (1792-1856) fue uno de los que, independientemente, descubrieron este tipo de geometría como alternativa a la de Euclides. Otros fueron Cari Friedrich Gauss (1777-1855), Ferdinand Schweickard y Janos Bolyai.

que la geometría lobachevskiana) en escalas considerablemente menos remotas que las cosmológicas, aunque

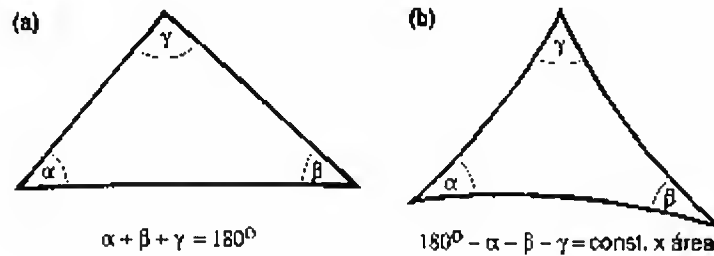


FIGURA V.1. (a) *Triángulo en un espacio euclidiano*, (b) *triángulo en un espacio lobachevskiano.*

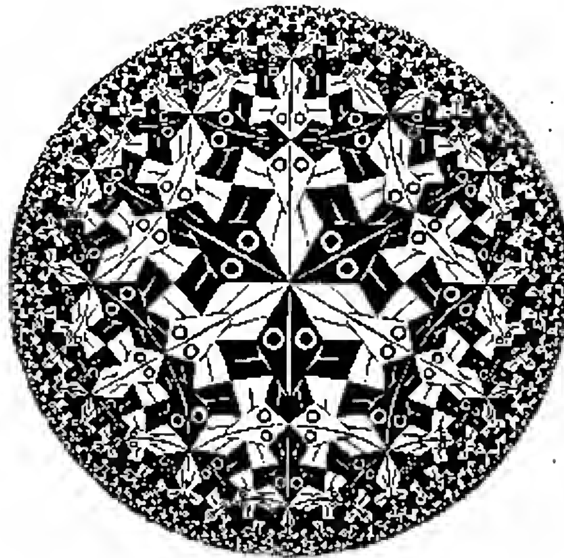


FIGURA V.2. *Representación, según Escher, del espacio de Lobachevsky.* (*Todos los peces negros deben considerarse congruentes*; *lo mismo debe hacerse con los peces blancos*.)

las desviaciones son todavía extraordinariamente pequeñas en las escalas ordinarias de nuestra experiencia directa.

El hecho de que la geometría euclidiana parezca tan precisa para reflejar la estructura del "espacio" de nuestro mundo nos ha engañado (o a nuestros predecesores) haciéndonos pensar que esta geometría es una necesidad lógica, o haciéndonos pensar que tenemos una intuición, innata *a priori*, de que la geometría euclidiana *debe* aplicarse al mundo en que vivimos. (Incluso el gran filósofo Emmanuel Kant afirmaba esto.) La verdadera ruptura con la geometría euclidiana sólo llegó con la teoría de la relatividad general de Einstein, propuesta muchos años después. El que la geometría euclidiana se aplique de forma tan precisa —aunque no suficientemente exacta— a la estructura de nuestro espacio físico, lejos de ser una necesidad lógica, es un *hecho observacional empírico*. La geometría euclidiana fue realmente, desde el principio, una teoría física SUPREMA. Lo era así además de ser un elemento elegante y lógico de la matemática pura.

En cierto sentido, esto no estaba tan alejado del punto de vista adoptado por Platón (*c*. 360 a.C.; esto es, unos cincuenta años antes de los *Elementos*, el famoso libro de geometría de Euclides). En opinión de Platón, los objetos de la geometría pura —líneas rectas, círculos, triángulos, planos, etc.— sólo se realizaban aproximadamente en el mundo de las cosas físicas reales. Los objetos matemáticamente precisos de la geometría pura no poblaban este mundo físico sino un mundo diferente: *el mundo ideal* de Platón de los conceptos matemáticos. El mundo de Platón consta no de objetos tangibles sino de "objetos matemáticos". Este mundo no nos es accesible del modo físico ordinario sino por la vía del *intelecto*. Nuestra mente entra en contacto con el mundo de Platón cada vez que contempla una verdad matemática, percibiéndola mediante el ejercicio del razonamiento y la intuición matemática. El mundo ideal se consideraba diferente y más perfecto que el mundo material de nuestra experiencia externa, pero tan real como éste. (Téngase en cuenta lo dicho en los capítulos III y IV, sobre la realidad platónica de los conceptos matemáticos.) Así, mientras que los objetos de la geometría euclidiana pura pueden ser estudiados por el pensamiento, y pueden derivarse de este modo muchas propiedades de este ideal, no hay necesidad de que el "imperfecto" mundo físico de la experiencia externa se ajuste exactamente a este ideal. Por una milagrosa intuición, y sobre la base de lo que debieron ser datos muy dispersos en ese tiempo, Platón parece haber previsto esto: por una parte, las matemáticas deben estudiarse y comprenderse por sí mismas, y no debemos pedir su aplicabilidad exacta a los objetos de la experiencia física; por otra parte, el funcionamiento del mundo externo real puede ser entendido finalmente sólo por virtud de las matemáticas exactas, lo que, en términos del mundo ideal de Platón, quiere decir "accesible por la vía del intelecto".

Platón fundó en Atenas la Academia destinada a fomentar tales ideas Entre la élite que surgió de sus miembros estaba Aristóteles, el filósofo más influyente y famoso que haya existido. Pero aquí nos interesaremos en otro miembro de la Academia (algo menos conocido que Aristóteles pero, en mi opinión, mucho mejor científico), uno de los más grandes pensadores de la Antigüedad: el matemático y astrónomo Eudoxo.

Existe un ingrediente profundo y sutil en la geometría euclidiana —en realidad el más esencial— y que hoy en día apenas lo consideramos como geometría. (Los matemáticos tenderán a llamar a este elemento "análisis" más que "geometría".) Éste constituía la introducción efectiva a los *números reales*. La geometría euclidiana trabaja con longitudes y ángulos. Para comprender esta geometría debemos estimar qué tipo de "números" son necesarios para describir esas longitudes y ángulos. La nueva idea central fue expuesta en el siglo IV a.C. por Eudoxo (*c*. 408-355 a.C.).[*] La geometría griega había pasado por una "crisis" debido al descubrimiento que los pitagóricos hicieron de que números como $\sqrt{2}$ (necesarios para expresar la relación entre la longitud de la diagonal de un cuadrado y su lado) no pueden expresarse como una fracción (*cfr*. capítulo III). Hubiera sido importante para los griegos el poder formular sus medidas (razones) geométricas en términos de (razones de) enteros para que las magnitudes geométricas pudieran ser estudiadas de acuerdo con las leyes de la aritmética. Básicamente, la idea de Eudoxo fue proporcionar un método para describir razones de longitudes (esto es, ¡números reales!) en términos de *enteros*. Él fue capaz de dar criterios, establecidos en términos de operaciones enteras, para decidir cuándo una razón es mayor que otra, o si las dos deben considerarse exactamente iguales.

[*] Eudoxo fue también quien dio origen a la teoría ÚTIL (de 2 000 años de duración) del movimiento planetario, posteriormente desarrollada con más detalle por Hiparco y Tolomeo, y conocida en consecuencia como sistema tolemaico.

La idea era aproximadamente la siguiente: si a, b, c y d son cuatro longitudes, entonces un criterio para verificar que la razón a/b es *mayor* que la razón c/d es que existan números enteros M y N tales que a sumada N veces consigo misma sea mayor que b sumada M veces consigo misma, al tiempo que d sumada M veces consigo misma sea mayor que c sumada N veces consigo misma.** Un criterio análogo puede utilizarse para verificar que a/b es *menor* que c/d. El criterio buscado para la *igualdad* $a/b = c/d$ es entonces simplemente que *ninguno* de esos otros dos criterios pueda satisfacerse.

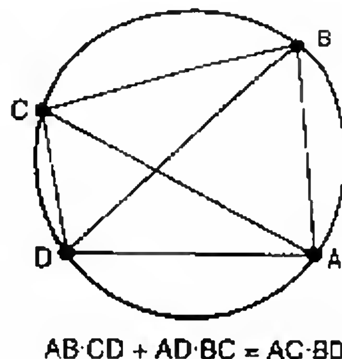


FIGURA V.3. *El teorema De Tolomeo.*

Una teoría matemática abstracta completamente exacta de los números reales no fue desarrollada hasta el siglo XIX, por matemáticos como Dedekind y Weierstrass. Pero sus métodos seguían realmente líneas muy similares a las que Eudoxo ya había descubierto unos veintidós siglos antes. No hay necesidad aquí de describir este moderno avance. Esta teoría moderna fue vagamente insinuada en el capítulo III, pero entonces preferí, para facilitar la presentación, basar el tratamiento de los números reales en las más conocidas expansiones decimales (éstas fueron introducidas por Stevin en 1585). Debe tenerse en cuenta que la notación decimal, aunque familiar para nosotros, era desconocida para los griegos.

Hay una diferencia importante, sin embargo, entre la propuesta de Eudoxo y las de Dedekind y Weierstrass. Los antiguos griegos pensaban en los números reales como cosas *dadas* —como razones de magnitudes geométricas— es decir, como propiedades del espacio "real". Era necesario para los griegos poder describir las magnitudes geométricas en términos aritméticos para poder razonar rigurosamente sobre ellas, y también sobre sus sumas y productos, ingredientes esenciales de muchos de los maravillosos teoremas geométricos de los antiguos. (En la fig. V.3 he dado, a modo de ilustración, el famoso *teorema de Tolomeo* —aunque él lo descubrió en una época muy posterior a Eudoxo— que relaciona las distancias entre cuatro puntos de una circunferencia, que ilustra muy bien cómo son necesarias ambas sumas y productos.) El criterio de Eudoxo se mostró extraordinariamente fructífero y, en particular, capacitó a los griegos para calcular rigurosamente áreas y volúmenes.

Sin embargo, para los matemáticos del siglo XIX —y, ciertamente. Para los de hoy— el papel de la geometría ha cambiado. Para los antiguos griegos, y para Eudoxo en particular, los números

** En notación moderna esto afirma la existencia de una fracción, a saber, M/N, tal que $a/b > M/N > c/d$. Siempre existirá tal fracción entre los dos números a/b y c/d con tal de que $a/b > c/d$, de modo que el criterio de Eudoxo se satisface efectivamente.

"reales" eran cosas que había que *extraer* de la geometría del espacio físico. Ahora preferimos concebir los números reales como lógicamente más primitivos que la geometría. Esto nos permite construir toda clase de tipos *diferentes* de geometría, *partiendo* para cada uno del concepto de número. (La idea clave fue la de geometría de coordenadas, introducida en el siglo XVII por Fermat y Descartes. Las coordenadas pueden utilizarse para *definir* otros tipos de geometría.) Cualquiera de estas "geometrías" debe ser lógicamente consistente, pero no es necesario que tenga que ver directamente con el espacio físico de nuestra experiencia. La geometría física que nos *parece* percibir es una *idealización* de la experiencia (*v.g.* dependiente de nuestras extrapolaciones a tamaños indefinidamente grandes o pequeños, *cfr.* capítulo III), pero los experimentos son ahora lo suficientemente precisos y debemos aceptar que nuestra geometría "experimentada" *difiere* realmente del ideal euclidiano, y es compatible con lo que, según la teoría de la relatividad general de Einstein, debería ser. Sin embargo, a pesar de los cambios que han tenido lugar en nuestra visión de la geometría del mundo físico, el concepto eudoxiano de número real, con veintitrés siglos de edad, ha permanecido sin mayores cambios y constituye un ingrediente tan esencial en la teoría de Einstein como en la de Euclides. En realidad, ha sido ingrediente fundamental de cualquier teoría física seria hasta nuestros días.

El libro quinto de los *Elementos* de Euclides era básicamente una exposición de la "teoría de las proporciones", descrita arriba, que introdujo Eudoxo. Ésta era profundamente importante para la obra en conjunto. En realidad, los *Elementos* en su totalidad, publicados por primera vez alrededor del año 300 a.C., deben conceptuarse como una de las obras de más profunda influencia de todos los tiempos. Ellos sientan el escenario de casi todo el pensamiento científico y matemático a partir de entonces. Sus métodos eran deductivos, partían de axiomas claramente enunciados que se suponían propiedades "evidentes por sí mismas" del espacio; de éstos se derivaban numerosas consecuencias, muchas de las cuales eran sorprendentes e importantes, y en absoluto evidentes. No hay duda de que la obra de Euclides tuvo una profunda significación para el desarrollo del pensamiento científico posterior.

El matemático más grande de la Antigüedad fue indudablemente Arquímedes (287-212 a.C.). Utilizando ingeniosamente la teoría de las proporciones de Eudoxo, calculó las áreas y volúmenes de muchas formas geométricas diferentes, como la esfera, u otras más complejas, entre ellas las parábolas y las espirales. Hoy utilizaríamos el cálculo integral para hacerlo, pero esto ocurría unos diecinueve siglos antes de la introducción del cálculo infinitesimal por Newton y Leibniz. (Podría decirse que una buena mitad —la mitad "integral"— del cálculo ya era conocida para Arquímedes.) El grado de rigor matemático que este sabio alcanzó en sus argumentos era impecable, incluso para las exigencias modernas. Sus escritos influyeron profundamente sobre muchos matemáticos y científicos posteriores, muy en especial Galileo y Newton. Arquímedes introdujo también la (¿SUPREMA?) teoría física de la estática (es decir, las leyes que gobiernan los cuerpos en equilibrio, como la ley de la palanca y las leyes de los cuerpos flotantes) y la desarrolló como ciencia deductiva, de un modo semejante a como Euclides había desarrollado la geometría del espacio y la de los cuerpos rígidos.

Un contemporáneo de Arquímedes a quien también debe mencionarse es Apolonio (*c.* 262-200 a.C.), geómetra de profunda intuición e ingenio, cuyo estudio de la teoría de las secciones cónicas (esto es, elipses, parábolas e hipérbolas) tuvo una influencia muy importante sobre Kepler y Newton. Precisamente estas figuras geométricas resultaron ser, de forma bastante notable, las que se necesitaban para describir las órbitas planetarias.

LA DINÁMICA DE GALILEO Y NEWTON

El profundo cambio que el siglo XVII aportó a la ciencia fue la comprensión del *movimiento*. Los antiguos griegos tenían una maravillosa comprensión de la estática —formas geométricas rígidas, o cuerpos en equilibrio (es decir, cuando todas las fuerzas están compensadas de modo que no hay movimiento)—, pero no tenían una buena concepción de las leyes que gobiernan los cuerpos que se *mueven*. Lo que les faltaba era una buena teoría de la *dinámica*, esto es, una teoría del modo en que la naturaleza controla el cambio de posición de los cuerpos de un instante al siguiente. Parte (pero no todas ni mucho menos) de las razones para esto era la ausencia de cualquier medio suficientemente preciso para medir el tiempo, es decir, de un "reloj" razonablemente bueno. Un reloj así es necesario para poder cronometrar exactamente los cambios en posición, y de este modo comprobar las velocidades y aceleraciones de los cuerpos. Por ello, la observación de Galileo, en 1583, de que un péndulo podía ser un medio confiable de medir el tiempo tuvo para él (y para el desarrollo de la ciencia moderna en general) una enorme importancia, puesto que permitió hacer un cronometraje preciso del movimiento.[4] Unos cincuenta y cinco años más tarde, con la publicación de los *Discorsi* de Galileo en 1638, nacería la nueva ciencia de la dinámica y empezaría a transformarse el antiguo misticismo en ciencia moderna.

Escogeré sólo *cuatro* de las ideas físicas más importantes que introdujo Galileo. La primera era que una fuerza que actúa sobre un cuerpo determina la *aceleración*, y no la velocidad. ¿Qué significan realmente los términos "aceleración" y "velocidad"? La *velocidad* de una partícula —o de un punto de algún cuerpo— es el ritmo de cambio, con respecto al tiempo, de la posición de dicho punto. Normalmente se toma la velocidad como una cantidad *vectorial*, lo que quiere decir que se debe considerar tanto su *dirección* como su magnitud (de lo contrario utilizamos el término "celeridad"; véase fig. V.4). La aceleración (de nuevo una cantidad vectorial) es el ritmo de cambio de esta velocidad con respecto al tiempo, de modo que la aceleración es realmente el *ritmo de cambio del ritmo de cambio* de la posición con respecto al tiempo. (Hubiera sido difícil para los antiguos entender esto, al faltarles los "relojes" y las ideas matemáticas adecuadas sobre los "ritmos de cambio".) Galileo comprobó que la fuerza que actúa sobre un cuerpo (en su caso, la fuerza de la gravedad) controla la aceleración de dicho cuerpo pero *no* controla directamente su velocidad, tal como los antiguos, por ejemplo Aristóteles, habían creído.

En particular, si no hay fuerza la velocidad es constante y, por lo tanto, en *ausencia* de fuerzas resultará un movimiento uniforme en línea recta (lo que constituye la primera ley de Newton). Los cuerpos en movimiento libre continúan uniformemente su camino, y no se necesita ninguna fuerza que mantenga su marcha. De hecho, una consecuencia de las leyes dinámicas que establecieron Galileo y Newton es que el movimiento rectilíneo uniforme es físicamente indistinguible del estado de reposo (es decir, de ausencia de movimiento): no existe modo de distinguir localmente el movimiento uniforme del estado de reposo. Galileo fue especialmente claro en este punto (incluso más claro que Newton) y

[4] Parece ser, sin embargo, que Galileo usó a menudo una clepsidra para medir el tiempo en sus observaciones, véase Barbour, 1989.

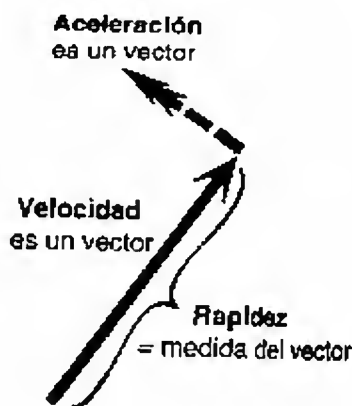


FIGURA V.4. *Velocidad, celeridad y aceleración.*

dio una descripción muy gráfica recurriendo al caso de un barco en el mar (*cfr.* Drake, 1953, pp. 186-187):

> Encerrémonos con un amigo en la cabina principal bajo la cubierta de un gran barco, llevando con nosotros moscas, mariposas y otros pequeños animales voladores. Llevemos un gran recipiente con agua y algún pez dentro; colguemos una botella que se vacíe gota a gota en alguna vasija que esté debajo de ella. Con el barco aún en reposo, observemos cuidadosamente cómo vuelan los pequeños animales con igual velocidad hacia todos los lados de la cabina. El pez nadará indistintamente en todas las direcciones; las gotas caerán en la vasija inferior... Cuando hayamos observado cuidadosamente todas estas cosas,... hagamos avanzar el barco con la velocidad que queramos, de forma que el movimiento sea uniforme y no haya oscilaciones en un sentido u otro. No descubriremos el menor cambio en ninguno de los efectos mencionados, ni podríamos decir a partir de ellos si el barco se mueve o permanece quieto... Las gotas caerán como antes en la vasija inferior sin desviarse hacia la popa, aunque el barco haya avanzado mucho mientras las gotas están en el aire. El pez nadará hacia la parte delantera de su recipiente sin mayor esfuerzo que hacia la parte trasera, y se dirigirá con la misma facilidad hacia un cebo colocado en cualquier parte del borde del recipiente. Finalmente, las mariposas y moscas continuarán su vuelo indistintamente hacia cualquier lado, y no sucederá que se concentren hacia la popa como si se cansaran de seguir el curso del barco, del que hubieran quedado separadas una gran distancia de haberse mantenido en el aire.

Este notable hecho, llamado *principio de relatividad galileana*, es realmente crucial para que tenga sentido dinámico el punto de vista *copernicano*. Nicolás Copérnico (1473-1543, y el antiguo astrónomo griego Aristarco, *c.* 310-230 a.C.), había presentado la imagen en la que el Sol permanece en reposo mientras que la Tierra, al mismo tiempo que gira sobre su propio eje, se mueve en una órbita en torno al Sol. ¿Por qué no somos conscientes de este movimiento, que sería de unos 100000 kilómetros por hora? Antes de que Galileo presentase su teoría dinámica, este hecho planteaba un verdadero y profundo enigma para el punto de vista copernicano. Si hubiera sido correcta la anterior visión "aristotélica" de la dinámica, en la que la *velocidad* real de un sistema en su movimiento a través del espacio afectaría a su comportamiento dinámico, entonces el movimiento de la Tierra sería en verdad muy directamente evidente para nosotros. La

relatividad galileana pone en claro como puede estar la Tierra en movimiento aunque su movimiento no sea algo que podamos percibir directamente.*

Nótese que, según la relatividad galileana, no se puede asociar ningún significado físico local al concepto de estar "en reposo". Esto tiene ya notables consecuencias para el modo de considerar el espacio y el tiempo. La imagen que intuitivamente tenemos sobre ellos es que el "espacio" constituye una especie de escenario en el que tienen lugar los sucesos físicos. Un objeto físico puede estar en un punto del espacio en un instante y en ese mismo punto o en otro en un instante posterior. Imaginemos que los puntos del espacio persisten de alguna manera entre un instante y el siguiente, de modo que tenga significado decir que el objeto ha cambiado o no su posición espacial. Ahora bien, la relatividad galileana nos dice que no hay significado absoluto para el "estado de reposo", así que no se puede asociar ningún significado "al mismo punto del espacio en dos instantes diferentes". ¿Qué punto del espacio euclidiano tridimensional de la experiencia física en un instante dado es el "mismo" punto de nuestro espacio euclidiano tridimensional en otro instante? No hay manera de decirlo. Parecería como si debiéramos tener un espacio euclidiano completamente *nuevo para cada* instante de tiempo. La manera como esto cobra sentido es considerar una imagen de la realidad física en un *espacio-tiempo tetradimensional* (véase la fig. V.5). Los espacios euclidianos tridimensionales correspondientes a los diferentes instantes de tiempo se consideran realmente como independientes uno de otro, pero todos estos espacios están unidos para formar la imagen completa de nuestro espacio-tiempo tetradimensional. Las trayectorias de las partículas que se mueven con movimiento rectilíneo uniforme se describen mediante líneas rectas (llamadas líneas de universo) en el espacio-tiempo. Volveremos más adelante a la cuestión del espacio-tiempo, y la relatividad del movimiento, en el contexto de la relatividad einsteiniana. Encontraremos que en este caso el argumento a favor de la tetradimensionalidad tiene una fuerza considerablemente mayor.

La tercera de estas grandes intuiciones de Galileo dio inicio de una comprensión de la *conservación de la energía*. Galileo estaba interesado principalmente en el movimiento de los objetos sometidos a la gravedad. Notó que si un cuerpo queda liberado a partir del reposo, entonces ya sea que caiga libremente, o que cuelgue de un péndulo de longitud arbitraria, o que se deslice por un plano inclinado, su velocidad de movimiento depende siempre *sólo* de la distancia que ha alcanzado por debajo del punto en que se ha soltado. Además su velocidad es siempre la justa para volver a la altura de la que partió. Como diríamos ahora, la

* Estrictamente hablando, esto se refiere sólo al movimiento de la Tierra en tanto que Pueda considerarse como aproximadamente *uniforme y*, en particular, sin rotación. El movimiento de rotación de la Tierra tiene realmente efectos dinámicos (relativamente pequeños) y detectables, siendo el más digno de mención la desviación de los vientos en sentidos diferentes en los hemisferios Norte y Sur. Galileo pensaba que esta falta de uniformidad era responsable de las mareas.

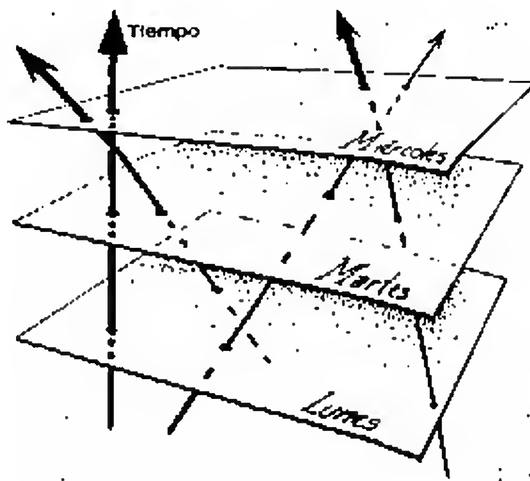


FIGURA V.5. *Espacio-tiempo galileano*: *las partículas en movimiento uniforme se representan como líneas rectas.*

energía almacenada en su altura sobre el suelo (energía potencial gravitatoria) puede transformarse en la energía de su movimiento (energía cinética que depende de la *celeridad* del cuerpo) y viceversa, pero la energía en conjunto no se gana ni se pierde.

La ley de la conservación de la energía es un principio físico muy importante. No es un requisito físico independiente, sino una *consecuencia* de las leyes dinámicas de Newton a las que llegaremos en breve. En el curso de los siglos hicieron formulaciones cada vez más generales de esta ley, Descartes, Huygens, Leibniz, Euler y Kelvin. Volveremos a ello después, en éste y en el capítulo VIL Resulta que, cuando se combina con el principio de relatividad de Galileo, la conservación de la energía da nuevas leyes de conservación de considerable importancia: conservación de *masa* y de *momento*. El momento de una partícula es el producto de su masa por su velocidad. Ejemplos conocidos de conservación del momento ocurren en los cohetes a reacción, en donde el incremento de momento hacia adelante del cohete compensa exactamente el momento hacia atrás de los gases expulsados (de menor masa pero, en compensación, mucho más rápidos). El retroceso de un fusil es también una manifestación de una conservación del momento. Otra consecuencia de las leyes de Newton es la conservación del *momento angular* que describe la persistencia de la rotación de un sistema. Tanto la rotación de la Tierra en torno a su eje como la de una pelota de tenis se mantienen en virtud de la conservación de su momento angular. Cada partícula de un cuerpo contribuye al momento angular del mismo, para el que la magnitud de la contribución de una partícula es el producto de su momento por su distancia perpendicular al centro. (En consecuencia, puede incrementarse la velocidad angular de un objeto que rota libremente si se le hace más compacto. Esto conduce al sorprendente, aunque bien conocido, molinete que suelen realizar los patinadores y trapecistas. El acto de recoger los brazos o las piernas, según sea el caso, provoca un incremento espontáneo de la velocidad de rotación, debido simplemente a la conservación del momento angular.) Veremos más adelante que masa, energía, momento y momento angular son conceptos importantes para nosotros.

Finalmente, recordaré al lector la profética intuición de Galileo de que, no habiendo fricción atmosférica, todos los cuerpos sometidos a la gravedad caen con la misma velocidad. (Quizá recuerde el lector la famosa anécdota de Galileo dejando caer simultáneamente varios objetos desde la torre inclinada de Pisa.) Tres siglos después, esta misma intuición condujo a Einstein a

generalizar el principio de relatividad a los sistemas de referencia acelerados, y proporcionó la piedra angular de su extraordinaria teoría de la relatividad general para la gravitación, como veremos cerca del final de este capítulo.

Sobre los impresionantes cimientos que dejó Galileo, Newton pudo levantar una catedral de soberana grandeza. Newton dio las tres leyes que gobiernan el comportamiento de los objetos materiales. La primera y segunda leyes eran en esencia las dadas por Galileo: si sobre un cuerpo no actúa fuerza alguna, éste continuará moviéndose uniformemente en línea recta; si una fuerza actúa sobre él, entonces el producto de su masa por su aceleración (es decir, el ritmo de cambio de su momento) será igual a dicha fuerza. Una de las intuiciones propias de Newton fue el darse cuenta de la necesidad de una tercera ley: la fuerza que un cuerpo *A* ejerce sobre un cuerpo *B* es exactamente igual y opuesta a la fuerza que el cuerpo *B* ejerce sobre el cuerpo *A* ("para toda acción existe una reacción igual y en sentido opuesto"). Esto proporciona el marco básico. El "universo newtoniano" consta de partículas que se mueven en un espacio que está sujeto a las leyes de la geometría euclidiana. Las aceleraciones de dichas partículas están determinadas por las fuerzas que actúan sobre ellas. La fuerza sobre cada partícula se obtiene sumando (con la *ley de suma vectorial*; véase la fig. V.6) todas y cada una de las contribuciones separadas a la fuerza sobre esa partícula debidas a todas las *demás* partículas. Para que el sistema esté bien definido se necesita alguna regla precisa que nos diga cuál sería la fuerza que aparece sobre la partícula *A* debida a otra partícula *B*. Normalmente se exige que esta

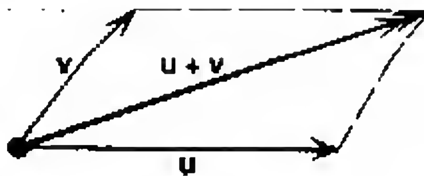


FIGURA V.6. *Ley del paralelogramo para la suma vectorial.*

fuerza actúe a lo largo de una línea recta entre *A* y *B* (véase la fig. V.7). Si la fuerza es de índole gravitatoria, entonces actúa atractivamente entre *A* y *B* y su intensidad es proporcional al producto de las dos masas y a la inversa del cuadrado de la distancia entre ellas: la *ley del inverso del cuadrado*. Para otros tipos de fuerza podría haber una dependencia de la posición distinta de ésta, y la fuerza podría depender de las partículas de acuerdo con alguna cualidad que posean distinta de sus masas.

El gran Johannes Kepler (1571-1630), contemporáneo de Galileo, había notado que las órbitas de los planetas en torno al Sol eran *elípticas* más que circulares (con el Sol situado siempre en un foco de la elipse, no en su centro) y formuló otras dos leyes que gobiernan los ritmos con que se describen las elipses. Newton pudo demostrar que las tres leyes de Kepler se deducen de su propio esquema general (con una ley de fuerza atractiva inversa del cuadrado). No sólo esto, sino que también obtuvo todo tipo de correcciones detalladas a las órbitas elípticas de Kepler, así como otros efectos, tales como la precesión de los equinoccios (un lento movimiento de la dirección del eje de rotación, que los griegos habían notado a lo largo de los siglos). Para lograr todo esto, Newton tuvo que desarrollar muchas técnicas matemáticas, además del cálculo diferencial. El éxito excepcional de sus esfuerzos debió mucho a sus supremas habilidades matemáticas y a su igualmente extraordinaria intuición física.

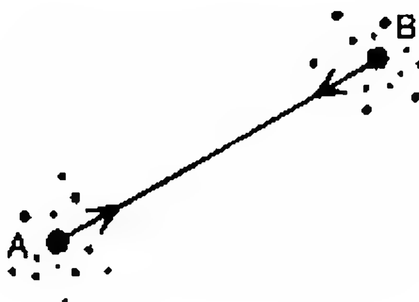


FIGURA V.7. *La fuerza entre Dos partículas se considera dirigida a lo largo de la línea recta que las une (y por la tercera ley de Newton, la fuerza sobre A debida a B es siempre igual y opuesta a la fuerza sobre B debida a A).*

EL MUNDO MECANICISTA DE LA DINÁMICA NEWTONIANA

Con una ley de fuerzas específica (como la ley de la inversa del cuadrado para la gravitación) el esquema newtoniano se traduce en un sistema de ecuaciones dinámicas preciso y determinado. Si se especifican las posiciones, velocidades y masas de las diversas partículas en un instante, entonces sus posiciones y velocidades (y sus masas, pues éstas se consideran *constantes)* están matemáticamente determinadas para todos los instantes posteriores. Esta forma de *determinismo,* satisfecha por el mundo de la mecánica newtoniana, tuvo (y aún tiene) una profunda influencia sobre el pensamiento filosófico. Tratemos de examinar un poco más de cerca la naturaleza de este determinismo newtoniano. ¿Qué puede decirnos sobre la cuestión del "libre albedrío"? ¿Podría existir la mente en un mundo estrictamente newtoniano? ¿Puede un mundo newtoniano dar cabida siquiera a las máquinas computadoras?

Tratemos de ser todo lo concretos que podamos sobre este modelo "newtoniano" del mundo. Podemos suponer, por ejemplo, que todas las partículas constituyentes de la materia se consideran como puntos materiales, esto es, sin ninguna extensión espacial. De igual modo podríamos considerarlas como bolas esféricas rígidas. En uno u otro caso tendremos que suponer que nos son conocidas las leyes dinámicas, como la ley de atracción inversa del cuadrado de la teoría gravitatoria de Newton. Conviene modelar también otras fuerzas de la naturaleza, tales como la *eléctrica* y la *magnética* (estudiadas por primera vez en detalle por William Gilbert en 1600), o las fuerzas *nucleares* fuertes que ahora sabemos son las que unen a las partículas (protones, neutrones) que forman el núcleo atómico. Las fuerzas eléctricas se parecen a las gravitatorias en que también satisfacen la ley de la inversa del cuadrado, pero aquí las partículas semejantes se *repelen* entre sí (en lugar de atraerse, como en el caso gravitatorio) y no son las masas de las partículas las que gobiernan la intensidad de las fuerzas eléctricas mutuas sino sus *cargas eléctricas.* Las fuerzas magnéticas son también del tipo de la "inversa del cuadrado" como las eléctricas,[*] pero las fuerzas nucleares tienen una dependencia de la distancia bastante diferente, siendo extremadamente grandes para las distancias muy cortas que separan las partículas dentro de los núcleos atómicos pero insignificantes a distancias mayores.

[*] La diferencia entre el caso eléctrico y el magnético es que no parece que existan "cargas magnéticas" aisladas (es decir, polos Norte y Sur) en la naturaleza, sino que las partículas magnéticas son lo que se denomina "dipolos", esto es, minúsculos imanes (con Polos Norte y Sur inseparables).

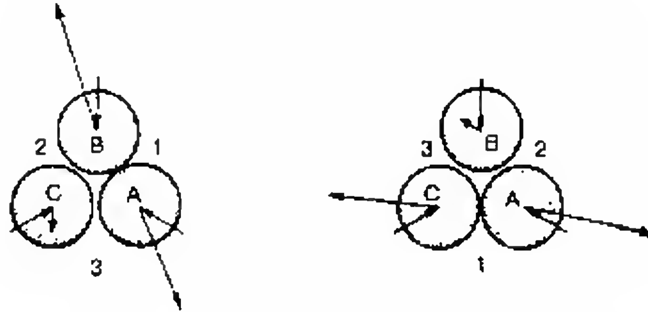


FIGURA V.8. *Colisión triple. El comportamiento final depende críticamente de qué partículas chocan primero, de modo que el resultado depende de manera discontinua de la situación inicial.*

Supongamos que adoptamos la imagen de la esfera rígida, exigiendo que cuando chocan entre sí dos de las esferas simplemente rebotan de forma perfectamente *elástica;* es decir, se separan de nuevo sin ninguna pérdida de energía (ni de momento total) como si fueran bolas de billar perfectas. También tenemos que especificar exactamente cómo son las *fuerzas* que actúan entre una bola y otra. Podemos suponer, en obsequio de la sencillez, que la fuerza que cada bola ejerce sobre cada una de las otras está dirigida a lo largo de la línea recta que une sus centros, y su magnitud es una determinada función de la longitud de esta línea. (Para la *gravitación* newtoniana esta suposición es automáticamente cierta, por un famoso teorema de Newton; y para otras leyes de fuerza puede imponerse como requisito constante.) Con tal de que las bolas choquen sólo en pares, y no ocurran colisiones triples o de orden mayor, entonces todo está bien definido y el resultado dependerá de manera continua del estado inicial (es decir, cambios suficientemente pequeños en el estado inicial conducen sólo a cambios pequeños en el resultado). Hay continuidad entre el comportamiento de las colisiones rasantes y el comportamiento de las bolas cuando apenas pasan una al lado de la otra. Hay, sin embargo, un problema en las colisiones triples o de orden mayor. Por ejemplo, si tres bolas A, B y C chocan al mismo tiempo, es diferente si consideramos que A y B chocan primero y C lo hace con B inmediatamente después, o si consideramos que son A y C las que chocan primero y B con A lo hacen inmediatamente después (véase fig. V.8). Nuestro modelo es *indeterminista* cuando ocurren colisiones triples exactas. Si lo fuéremos, podemos simplemente *excluir* las colisiones triples o de orden como "infinitamente improbables". Esto proporciona un esquema razonablemente coherente, pero el problema potencial de las colisiones triples significa que el comportamiento resultante puede *no* depender de forma continua del estado inicial.

Esto no es totalmente satisfactorio, y podemos preferir un modelo formado por partículas *puntuales.* Pero para evitar ciertas dificultades teóricas planteadas por este modelo (fuerzas infinitas y energías infinitas cuando las partículas llegan a coincidir) debemos hacer otras suposiciones, como la de que las fuerzas entre las partículas se hacen siempre fuertemente repulsivas a cortas distancias. De esta manera podemos asegurar que un par de partículas nunca colisionará realmente. (Esto también nos permite evitar el problema de cómo se supone que se *comportan* las partículas cuando chocan entre sí.) Sin embargo, para dar una idea más clara, prefiero plantear lo que sigue partiendo siempre de esferas rígidas. Al parecer, esta idea de las "bolas de billar" es esencialmente el modelo de *realidad* que tiene mucha gente.

Ahora (pasando por alto el problema de las colisiones múltiples), la imagen de la bola de billar newtoniana[5] de la realidad constituye efectivamente un modelo *determinista.* Debemos tomar la palabra "determinista" en el sentido de que el comportamiento físico está matemáticamente determinado en su totalidad para cualquier instante futuro (o pasado) por las posiciones y velocidades de todas las pelotas (supuestas en número finito, pongámoslo así, para evitar problemas) en *un* instante cualquiera. Parece, entonces, que en un mundo de bolas de billar no hay lugar para una "mente" que influya en el comportamiento de las cosas materiales mediante la acción de su "libre albedrío". Si creyéramos en el "libre albedrío" al parecer nos veríamos obligados a dudar de que nuestro mundo real estuviera formado de ese modo.

La controvertida cuestión del "libre albedrío" ronda en el trasfondo de este libro —aunque, para la mayor parte de lo que tendré que decir, quedará sólo en el trasfondo—. Tendrá un papel específico, aunque menor, que desempeñar más adelante en este mismo capítulo (en relación con el tema del envío de señales más rápidas que la luz en relatividad). La cuestión del libre albedrío se trata directamente en el capítulo X, y allí el lector quedará sin duda desilusionado por lo que tengo que aportar. Creo que existe aquí un auténtico problema, y no sólo uno supuesto, pero es profundo y difícil de formular adecuadamente. El problema del *determinismo* en la teoría física es importante pero creo que no es sino una parte del asunto. El mundo podría ser, por ejemplo, determinista pero *no computable.* En tal caso, el futuro podría estar determinado por el presente de un modo que *en principio* es no calculable. En el capítulo X trataré de presentar argumentos para demostrar que la acción de nuestra mente consciente es en realidad no algorítmica (es decir, no computable). En consecuencia, el libre albedrío del que nosotros mismos nos creemos dotados tendría que estar íntimamente ligado a algún ingrediente no computable en las leyes que gobiernan el mundo en que vivimos. Una cuestión interesante —ya sea que aceptemos o no este punto de vista respecto al libre albedrío— es la de si una teoría física dada (corno la de Newton) es en realidad *computable,* y no ya sólo si es determinista. La computabilidad es una cuestión diferente del determinismo, y el hecho de que *sea* una cuestión diferente es algo que trato de poner de relieve en este libro.

¿ES COMPUTABLE LA VIDA EN EL MUNDO DE LAS BOLAS DE BILLAR?

Permítaseme ilustrar primero, con un ejemplo absurdamente artificial, el hecho de que computabilidad y determinismo *son* diferentes, exhibiendo un "modelo de universo de juguete" que es determinista pero no computable. Describamos el "estado" del sistema en cualquier "instante" mediante un par de números naturales (m,n). Sea T_u una máquina universal de Turing bien determinada, por ejemplo la definida concretamente en el capítulo II . Para decidir cuál es el estado de este universo en el próximo "instante" debemos preguntar si la acción de T_u sobre m llegará a detenerse o no (es decir, si $T_u(m) \neq \square$ o $T_u(m) = \square$ en la notación del capítulo II. Si se detiene, el estado en este próximo "instante" es $(m + 1, n)$. Si no se detiene, será $(n + 1, n)$. Vimos en el capítulo II que no existe algoritmo para el problema de la detención de la máquina de Turing. De ello se sigue que no puede haber algoritmo para predecir el "futuro" en este universo modelo, a pesar de ser completamente determinista.[6] Por supuesto, este no es un modelo

[5] El nombre de Newton se asocia a este modelo —y en realidad a la mecánica "newtoniana" en general— simplemente como una conveniente *etiqueta.* Las propias opiniones de Newton respecto a la *verdadera* naturaleza del mundo físico parecen haber sido mucho menos dogmáticas y más sutiles que esto. (Al parecer, la persona que promovió con más fuerza este modelo "newtoniano" fue R. G. Boscovich, 1711-1787.)

[6] Raphael Sorkin me señala que en un cierto sentido la evolución de este modelo de juguete en particular puede ser "computada"

para tomar en serio, pero muestra que *existe* un problema. Podemos pedir a *cualquier* teoría física determinista que sea o no computable. De hecho, ¿es computable el mundo de las bolas de billar newtonianas?

El tema de la computabilidad física depende en parte del tipo de pregunta que nos propongamos plantearle al sistema. Puedo imaginar cierto número de preguntas que podrían plantearse y para las que mi conjetura sería que *no* son computables (esto es, que no es asunto algorítmico verificar la respuesta) en un modelo de bolas de billar newtonianas. Una de estas cuestiones podría ser: ¿choca alguna vez la bola *A* con la bola B? La idea es que se nos darían, como *datos iniciales,* las posiciones y velocidades de todas las bolas en cierto instante de tiempo (f = 0) y el problema consiste en calcular a partir de estos datos si las bolas *A* y *B* chocarán o no alguna vez en cualquier instante posterior ($t > 0$). Para hacer el problema más concreto (aunque no particularmente realista) podemos suponer que todas las bolas son del mismo radio y la misma masa y que hay, por ejemplo, una ley de fuerzas del tipo de la inversa del cuadrado actuando entre cada par de bolas. Una razón para conjeturar que esta pregunta concreta no es de las que pueden resolverse algorítmicamente es que el modelo es en cierto modo semejante a un "modelo de bolas de billar para una computación" que fue introducido por Edward Fredkin y Tommaso Toffoli (1982). En su modelo (en lugar de tener una ley de fuerzas de la inversa del cuadrado) las bolas están limitadas por varias "paredes", pero rebotan elásticamente una sobre otra de modo semejante a como lo hacen las bolas newtonianas que acabo de describir (véase fig. V.9) En el modelo de Fredkin-Toffoli todas las operaciones lógicas básicas de un ordenador pueden ejecutarse mediante las bolas. Puede imitarse cualquier computación de una máquina de Turing: la elección particular de la máquina de Turing T_n define la configuración de las "paredes", etc., de la máquina de Fredkin-Toffoli; una vez hecho esto, un estado inicial de bolas codifica la información de la cinta de *input,* y la cinta de *output* de la máquina de Turing queda codificada en el estado final de las bolas. Así puede plantearse, en particular, la pregunta: ¿se parará alguna vez tal o cual máquina computadora de Turing? "Parada' puede entenderse como que la bola *A* choque finalmente con la bola *B.* El hecho de que esta cuestión no pueda ser resuelta algorítmicamente *indica* al menos que la pregunta newtoniana "¿choca alguna vez la bola *A* con la bola *B?",* que planteé inicialmente, tampoco podrá responderse algorítmicamente.

En realidad, el problema newtoniano es mucho más complicado que el desarrollado por Fredkin y Toffoli. Estos pudieron especificar los estados de su modelo conforme a parámetros *discretos* (es decir, mediante enunciados "si o no" como "o la bola está en el canal o no lo está"). Pero en el problema newtoniano completo las posiciones y velocidades de las bolas tienen que especificarse con precisión infinita según coordenadas que son *números reales,* y no en forma discreta.

---

en un modo que no es del todo distinto del que utilizan, por ejemplo, los modelos newtonianos. Consideremos una sucesión de cómputos C_1, C_2, C_3,... que nos permite computar el comportamiento de nuestro sistema, tan lejano en el tiempo como queramos, sin límite alguno, y con una exactitud mayor. En el presente ejemplo, podemos conseguir esto definiendo C_N como el N-ésimo Paso de la máquina de Turing $T_u(m)$, *y* "considerando" $T_u(m)$ = □ si la acción de la máquina continúa en ese paso. Sin embargo, no sería difícil modificar nuestro modelo de juguete para que venciera a un "cómputo" como éste, sin embargo, al introducir una evolución que implica, en lugar de *Tû (m)* = □, enunciados doblemente cuantificados, como "T(q) se detiene para toda *q"*. (El problema sin resolver de que hay infinitas parejas de primos es decir dos números nones consecutivos que sean primos, es un buen ejemplo de un enunciado así.)

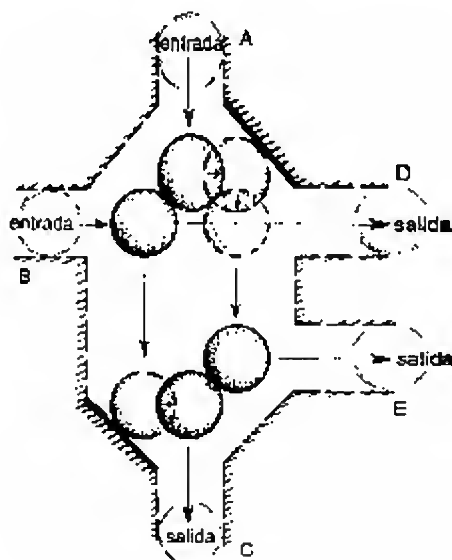


FIGURA V.9. *Un conmutador (sugerido por A. Resler) en la computadora de bolas de billar de Fredkin-Toffoli. Si un bola entra por* B, *otra saldrá a continuación por* D *o por E dependiendo de si otra bola entra por* A, *en donde las entradas por* A *y* B *se suponen simultáneas.*

De este modo, nos enfrentamos otra vez con todos los problemas que tuvimos que considerar cuando en el capítulo IV nos ocupábamos de la cuestión de si el conjunto de Mandelbrot es o no recursivo. ¿Qué *significa* "computable" cuando se admiten como datos de entrada y salida parámetros que varían de forma continua?[7] Por el momento, el problema puede atenuarse suponiendo que todas las coordenadas de posición y velocidad iniciales vienen dadas Por números *racionales* (aunque no podemos esperar que tales coordenadas sigan siendo racionales para posteriores valores racionales del tiempo *t).* Recordemos que un número racional es un cociente entre dos enteros; por consiguiente está definido en términos discretos finitos. Utilizando números racionales podemos aproximar, tanto como queramos, cualesquiera conjuntos de datos iniciales que hayamos decidido examinar. No es del todo descabellado conjeturar que, con datos iniciales racionales, pueda no existir algoritmo para decidir si finalmente chocarán las bolas *A* y *B.*

Sin embargo, esto no es realmente lo que queremos decir al afirmar que "el mundo de las bolas de billar newtonianas no es computable". El modelo particular que he estado comparando con nuestro mundo de bolas de billar newtonianas, a saber, la "computadora de bolas de billar" de Fredkin-Toffoli, actúa realmente de acuerdo con un cálculo. Éste, después de todo, era el punto esencial de la idea de Fredkin-Toffoli, que su modelo se comportara como una computadora (universal). El tipo de problema que trato de plantear es si es concebible que un cerebro humano pueda, aprovechando algunas leyes físicas "no computables", "superar" en algún sentido a una máquina de Turing. De nada sirve tratar de aprovechar algo como:

"Si la bola *A* nunca choca con la bola *B* entonces la respuesta a su problema es *no.* "

[7] Como se sugirió en el capítulo IV (nota 10), la nueva teoría de Blumb-Shub-Smale (1989) quizá proporcione un modo de resolver algunos de estos puntos de una forma matemáticamente más aceptable.

¡Acaso habría que esperar indefinidamente para asegurar que las bolas en cuestión no chocan nunca! Este es, por supuesto, el modo en que se *comporta* una máquina de Turing.

Parece que, en efecto, hay claros indicios de que, en un sentido apropiado, el mundo de las bolas de billar newtonianas *es* computable (al menos si prescindimos del problema de los choques múltiples). Normalmente trataríamos de computar este comportamiento haciendo algunas aproximaciones. Podríamos imaginar que se especifica que los centros de las bolas están en una malla de puntos, en donde los puntos nodales de la malla son precisamente aquellos que miden las coordenadas (en centésimas de unidad, por ejemplo). El tiempo se considera también "discreto": todos los instantes permitidos son múltiplos de alguna pequeña unidad (denotada por Δt , por ejemplo). Esto plantea algunas posibilidades discretas para las "velocidades" (diferencias en los valores de la posición de los nodos en dos instantes permitidos sucesivos divididas por Δt). Las aproximaciones apropiadas para las aceleraciones se calculan utilizando la ley de fuerzas, y estas aceleraciones se utilizan para obtener las "velocidades" a partir de las que se calculan con el grado de aproximación requerido, las nuevas posiciones de los nodos en el próximo instante permitido. El cálculo continúa durante tantos intervalos temporales como sea posible mientras se mantenga la precisión deseada. Quizá no se puedan computar muchos instantes antes de que se pierda toda precisión. El método consiste entonces en empezar con una malla espacial considerablemente más fina y una división algo más fina de los instantes de tiempo permitidos. Esto permite alcanzar una mayor precisión y el cálculo puede llevarse más lejos que en el caso anterior antes de que se pierda la precisión. Con una malla espacial aún más fina, y una división más sutil de los intervalos de tiempo, la precisión puede aumentarse aún más y llevar el cálculo aún más lejos. De esta forma, el mundo de las bolas de billar newtonianas puede computarse tan concienzudamente como queramos (pasando por alto las colisiones múltiples) —y, en este sentido, podemos decir que el mundo newtoniano es realmente computable.

Existe un sentido, no obstante, en el que este mundo es "no computable" *en la práctica.* Esto surge del hecho de que la precisión con que pueden *conocerse* los datos iniciales es siempre limitada. De hecho, existe una "inestabilidad" muy considerable inherente a este tipo de problemas. Un pequeñísimo cambio en los datos iniciales puede dar lugar rápidamente a un cambio absolutamente enorme en el comportamiento resultante. (Cualquiera que haya tratado de embuchacar una bola de billar americano, o *pool,* golpeándola con una bola intermedia que, a su vez, debe ser golpeada antes, sabrá lo que quiero decir.) Esto es particularmente evidente cuando se trata de colisiones (sucesivas), pero tales inestabilidades en el comportamiento pueden ocurrir también con la acción a distancia gravitatoria de Newton (con más de dos cuerpos). A menudo se utiliza el término "caos", o "comportamiento caótico", para este tipo de inestabilidad. El comportamiento caótico es importante, por ejemplo, en relación con el clima. Aunque se conocen las ecuaciones newtonianas que gobiernan los elementos, las predicciones del tiempo a largo plazo son muy poco confiables.

Este no es en absoluto el tipo de "no computabilidad" que pueda "aprovecharse" de forma alguna. Consiste simplemente en que, puesto que hay un límite a la precisión con que puede conocerse el estado inicial, el estado futuro no puede ser computado confiablemente a partir del inicial. En efecto, se ha introducido un *elemento aleatorio* en el comportamiento futuro, pero esto es todo. Si el cerebro humano realmente recurre a los elementos no computable.; *útiles* que existen en las leyes físicas, éstos deben ser de un tipo completamente diferente, de un carácter mucho más positivo. Consiguientemente, no me referiré a este tipo de comportamiento "caótico"

como "no computabilidad". Prefiero utilizar el término "impredecibilidad". La impredecibilidad es un fenómeno muy general en el tipo de leyes deterministas que aparecen en la física (clásica), como pronto veremos. La impredecibilidad es algo que ciertamente, más que "aprovechar", nos gustaría *minimizar* al construir máquina pensante.

Para ocuparnos de las cuestiones de computabilidad e impredecibilidad será conveniente adoptar un punto de vista más amplio que antes respecto a las leyes físicas. Esto nos facilitará el considerar no sólo el esquema de la mecánica newtoniana sino también las teorías posteriores que han venido a reemplazarla. Tendremos que echar una ojeada a la notable formulación *hamiltoniana* de la mecánica.

LA MECÁNICA HAMILTONIANA

Los éxitos de la mecánica newtoniana se derivaban no sólo de su extraordinaria aplicabilidad al mundo físico sino también de la riqueza de la teoría matemática a que dio lugar. Resulta notable que *todas* las teorías SUPREMAS de la naturaleza han resultado ser asaz fértiles como fuente de ideas matemáticas. Hay un misterio profundo y bello en el hecho de que estas teorías tan precisas sean también extraordinariamente fructíferas como simples *matemáticas.* Sin duda esto nos dice algo profundo sobre la conexión entre el mundo real de nuestra experiencia física y el mundo platónico de las matemáticas. (Intentaré abordar este tema más tarde, en el capítulo X.) La mecánica newtoniana ocupa tal vez un lugar supremo a este respecto puesto que su nacimiento dio lugar al cálculo infinitesimal. Además, el esquema newtoniano propiamente dicho ha dado lugar a un notable cuerpo de ideas matemáticas conocido como *mecánica clásica.* Los nombres de muchos de los grandes matemáticos de los siglos XVIII y XIX están asociados a este desarrollo: Euler, Lagrange, Laplace, Liouville, Poisson, Jacobi, Ostrogradski, Hamilton. Lo que se conoce como "teoría hamiltoniana"[8] resume gran parte de esta obra. Para nuestros propósitos, bastará una muestra de ella. El polifacético y original matemático irlandés William Rowan Hamilton (1805-1865) —a quien también se deben los circuitos hamiltonianos tratados, — había desarrollado esta forma de la teoría de una manera que realzaba la analogía con la propagación de las ondas. Esta idea de la relación entre las ondas y las partículas —y la forma de las propias ecuaciones de Hamilton— fue muy importante para el desarrollo posterior de la *mecánica cuántica.* Volveré a este aspecto de las cosas en el próximo capítulo.

Lo novedoso del esquema hamiltoniano reside en las "variables" que se utilizan en la descripción de un sistema físico. Hasta ahora, las *posiciones* de las partículas se consideraban como primarias, siendo las velocidades simplemente los ritmos de cambio de las posiciones respecto del tiempo. Recuérdese que en la especificación del estado inicial de un sistema newtoniano necesitábamos las posiciones *y* las velocidades de todas las partículas para determinar el comportamiento subsiguiente. Con la formulación hamiltoniana, más que las velocidades, debemos seleccionar los *momentos* de las partículas. (Señalamos anteriormente que el momento de una partícula es simplemente el producto de su velocidad por su masa.) Esto podría parecer en

[8] Las actuales ecuaciones de Hamilton, aunque quizá no su punto de vista particular, eran ya conocidas para el gran matemático Ítalo-francés Joseph C. Lagrange (1736-1813) unos 24 años antes que Hamilton. Anteriormente fue igualmente importante la formulación de la mecánica en términos de las *ecuaciones de Euler-Lagrange,* según las cuales las ecuaciones de Newton pueden verse como deducidas de un principio superior: el *principio de acción estacionaria* (P. L. M. de Maupertuis). Además de su gran importancia teórica. las ecuaciones de Euler-Lagrange proporcionan métodos de cálculo de considerable potencia y valor práctico.

sí mismo un cambio pequeño, pero lo importante es que la posición y el momento de cada partícula van a ser tratados como si fueran variables *independientes,* más o menos en pie de igualdad. De este modo se "pretende", en primer lugar, que los momentos de las diversas partículas no tengan nada que ver con los ritmos de cambio de sus variables de posición respectivas sino que son simplemente un conjunto separado de variables, de modo que podemos imaginar que "podrían" haber sido completamente independientes de los movimientos de posición. En la formulación hamiltoniana tenemos ahora *dos* conjuntos de ecuaciones. Uno de éstos nos dice cómo cambian con el tiempo los *momentos* de las diversas partículas, y el otro nos dice cómo cambian con el tiempo las *posiciones.* En ambos casos los ritmos de cambio están determinados por las diversas posiciones y momentos *en* dicho instante.

Hablando en términos generales, el primer conjunto de ecuaciones de Hamilton establece la crucial segunda ley de movimiento de Newton (Fuerza = ritmo de cambio del momento) mientras que el segundo conjunto de ecuaciones nos dice qué *son* realmente los momentos en función de las velocidades (en efecto, ritmo de cambio de la posición = momento ÷ masa). Recordemos que las leyes de movimiento de Galileo-Newton venían descritas en función de aceleraciones, es decir, ritmos de cambio de ritmos de cambio de posición (esto es, ecuaciones de "segundo orden"). Ahora sólo necesitamos hablar de ritmos de cambio de objetos (ecuaciones de "primer orden") y no de ritmos de cambio de ritmos de cambio de objetos. Todas estas ecuaciones se derivan simplemente de una cantidad importante: la *función hamiltoniana* o *hamiltoniano,* a secas, *H,* que es la expresión para la *energía total* del sistema en función de todas las variables de posición y momento.

La formulación hamiltoniana proporciona una descripción muy elegante y simétrica de la mecánica. Sólo para ver qué aspecto tiene, escribiremos aquí las ecuaciones aun cuando muchos lectores no estén familiarizados con las notaciones del cálculo que se requieren para una comprensión completa, que no serán necesarias aquí. Todo lo que realmente tenemos que comprender, en lo que se refiere al cálculo, es que el "punto" que aparece en el primer miembro de cada ecuación representa el *ritmo de cambio respecto del tiempo* (del momento, en el primer caso, y de la posición, en el segundo):

$$p_i = -\frac{\partial H}{\partial x_i}, \quad x_i = -\frac{\partial H}{\partial p_i},$$

Aquí el índice *i* se utiliza simplemente para distinguir las diferentes coordenadas de momento p_1 p_2, p_3, p_4,... y las de posición x_1, x_2, x_3, x_4,... para *n* partículas sin ligaduras tendremos 3n coordenadas de momento y 3n coordenadas de posición (una para cada una de las tres direcciones independientes del espacio). El símbolo ∂ denota la "derivada parcial" ("tomar derivadas mientras se mantienen constantes todas las demás variables"), y *H* es la función hamiltoniana descrita arriba. (Si usted no sabe nada sobre "diferenciación" no se preocupe. Piense sólo en los segundos miembros de estas ecuaciones como expresiones matemáticas perfectamente bien definidas escritas en términos de las x_i, y las p_i.)

Las coordenadas x_1, x_2,... y p_1, p_2,.... pueden ser cosas más generales que las simples coordenadas cartesianas de las partículas (esto es, las x_i siendo distancias medidas en tres direcciones diferentes que forman ángulos rectos). Algunas de las coordenadas x_i podrían ser *ángulos,* por ejemplo (en cuyo caso las correspondientes p_i serían momentos *angulares,* en lugar de momentos), o alguna otra magnitud general cualquiera. Curiosamente, las ecuaciones hamiltonianas aún mantienen exactamente la misma forma. De hecho, si elegimos

apropiadamente *H,* las ecuaciones de Hamilton siguen siendo verdaderas para *cualquier* sistema de ecuaciones clásicas, no sólo para las ecuaciones de Newton. Este será el caso, en particular, para la teoría de Maxwell-Lorentz que consideraremos dentro de poco. Las ecuaciones de Hamilton siguen siendo válidas también para la relatividad especial. Incluso la relatividad general, si se pone el cuidado debido, puede incluirse en el marco hamiltoniano. Además, como veremos más adelante con la ecuación de Schrödinger, este marco hamiltoniano proporciona el punto de partida para las ecuaciones de la mecánica cuántica. Semejante unidad formal en la estructura de las ecuaciones dinámicas, a pesar de todos los cambios revolucionarios que han ocurrido en las teorías físicas durante más o menos los últimos cien años, es algo verdaderamente notable.

ESPACIO DE FASES

La forma de las ecuaciones hamiltonianas nos permite "imaginar" de una manera muy clara y general la evolución de un sistema físico. Tratemos de imaginar un "espacio" de un gran número de dimensiones, una para cada una de las coordenadas x_1, x_2,..., p_1, p_2,... (Los espacios matemáticos tienen con frecuencia muchas más de tres dimensiones.) Este espacio se llama *espacio de fases* (véase fig. V.10). Para *n* partículas sin ligaduras, éste será un espacio de 6n dimensiones (tres coordenadas de posición y tres coordenadas de momento por cada partícula). La lectora o el lector podrán lamentarse de que incluso para una *sola* partícula, estas dimensiones son ya el doble de las dimensiones que normalmente se imaginan. El secreto está en no dejarse asustar por esto. Aunque seis dimensiones son realmente más dimensiones de las que podemos imaginar fácilmente (!) no sería tampoco de mucha utilidad que pudiéramos representarlas efectivamente. Sólo para una habitación llena de moléculas de aire, el número de dimensiones del espacio de fases podría ser algo como

10 000 000 000 000 000 000 000 000 000

No hay muchas esperanzas de hacerse una idea precisa de un espacio tan grande. Por ello, el secreto está en no intentarlo siquiera —aun en el caso del espacio de fases para una sola partícula—. Pensemos simplemente en un vago tipo de región tridimensional (o incluso solamente bidimensional). Echemos otra ojeada a la fig. V. 10. Eso bastará.

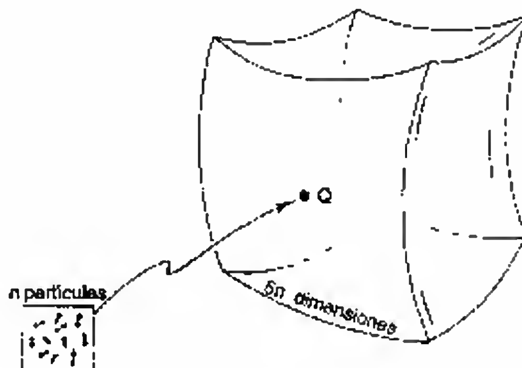


FIGURA V. 10. *Espacio de fases. Un simple punto Q del espacio de fases representa el estado global de algún sistema físico, incluyendo los movimientos instantáneos de cada una de sus partes.*

Ahora bien, ¿cómo vamos a imaginar las ecuaciones de Hamilton en términos del espacio de fases? En primer lugar, tenemos que considerar lo que realmente representa *un punto Q* en el espacio de fases. Corresponde a un conjunto concreto de valores para todas las coordenadas de posición x_1, x_2,... y para todas las coordenadas de momento p_1, p_2,... Es decir, Q representa nuestro *sistema físico completo,* con un estado de movimiento particular especificado para cada una de sus partículas simples constituyentes. Las ecuaciones de Hamilton nos dicen cuáles son los ritmos de cambio de todas estas coordenadas, una vez que sabemos sus valores presentes; es decir, gobiernan el comportamiento de todas y cada una de las partículas. Traducido en lenguaje del espacio de fases, las ecuaciones nos dicen cómo debe moverse un simple punto Q del espacio de fases, dada la localización actual de Q en dicho espacio. Así, en cada punto del espacio de fases tenemos una pequeña flecha —más correctamente, un *vector*— que nos dice cómo se está moviendo Q, para describir la evolución en el tiempo de nuestro sistema completo. La disposición global de flechas constituye lo que se conoce como un "campo vectorial" (fig. V.11). Por consiguiente, las ecuaciones de Hamilton definen un campo vectorial en el espacio de fases.

Veamos cómo debe interpretarse el determinismo físico en términos del espacio de fases. Para datos iniciales en el instante $t = 0$ tendremos un conjunto particular de valores especificados para todas las coordenadas de posición y momento; es decir, tendremos una elección particular del punto Q en el espacio de fases. Para hallar la evolución del sistema en el tiempo seguimos simplemente las flechas. Así, la evolución global de nuestro sistema con el tiempo —no importa cuan complicado pueda ser este sistema— se describe en el espacio de fases como un simple punto que se mueve siguiendo las flechas particulares que encuentra. Podemos pensar que las flechas indican la "velocidad" de nuestro punto Q en el espacio de fases. Si la flecha es "larga", Q se mueve rápidamente en su dirección, pero si la flecha es "corta" el movimiento de Q se hace más lento. Para ver lo que está haciendo nuestro sistema físico en el instante í, miramos simplemente hacia dónde se ha movido Q en ese momento, siguiendo las flechas de esta manera. Evidentemente este es un procedimiento determinista. La forma en que se mueve Q está completamente determinada por el campo vectorial hamiltoniano.

¿Qué sucede con la computabilidad? Si empezamos en un punto computable en el espacio de fases (esto es, en un punto en el que todas sus coordenadas de posición y momento son números computables, *cfr*-capítulo III) y esperamos a un tiempo computable t ¿terminaremos necesariamente en un punto que puede ser computado a partir de t y los valores de las coordenadas del punto de partida?

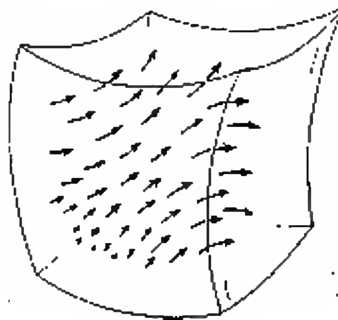


FIGURA V.11. *Campo de vectores del espacio de fases que representa la evolución temporal según las ecuaciones de Hamilton.*

La respuesta dependerá sin duda de la elección de la función hamiltoniana *H.* De hecho, habrá *constantes físicas* que aparecen en *H,* como son la constante gravitatoria de Newton o la velocidad de la luz —cuyos valores exactos dependerán del sistema de unidades elegido, aunque otras podrían ser puros números— y sería necesario asegurar que estas constantes son *números computables* para tener esperanzas de obtener una respuesta afirmativa. Si suponemos que éste *es* el caso, entonces mi *conjetura* sería que, para los hamiltonianos que normalmente se encuentran en física, la respuesta sería afirmativa. Sin embargo, esto *es* simplemente una conjetura y confío en que la pregunta, por el interés que entraña, es una cuestión que será examinada más a fondo en el futuro.

Por otra parte, tengo la impresión de que, por razones análogas a las que planteé hace poco en relación con el mundo de las bolas de billar, éste no es el resultado más importante. Se requeriría una *precisión infinita* para las coordenadas de un punto del espacio de fases —es decir, *todas* las cifras decimales— para que tuviera sentido decir que el punto es no computable. (Un número descrito con decimales *finitos* es siempre computable.) Una porción finita de una expansión decimal de un número no nos dice nada sobre la computabilidad de la expansión completa de dicho número. Pero todas las medidas físicas tienen un límite definido a la precisión con que pueden ser realizadas y sólo pueden dar información sobre un número finito de cifras decimales. ¿Anula esto el concepto global de "número computable" cuando se aplica a medidas físicas?

En realidad, un dispositivo que pudiera, de cualquier modo *útil,* sacar provecho de un (hipotético) elemento no computable en las leyes físicas probablemente no tendría que depender del hecho de hacer medidas de precisión ilimitada. Pero puede ser que esté adoptando aquí una postura demasiado estricta. Supongamos que tenemos un dispositivo físico que, por razones teóricas conocidas, imita algún interesante proceso matemático no algorítmico. El comportamiento exacto del dispositivo, de poder comprobarse exactamente, daría entonces las respuestas correctas a una serie de interesantes preguntas matemáticas del tipo sí/no para las que no puede haber un algoritmo (como los considerados en el capítulo IV). Cualquier algoritmo *dado* fallaría en alguna etapa, y en *dicha* etapa, el dispositivo nos daría algo nuevo. El dispositivo podría implicar el examen de algún parámetro físico con una precisión cada vez mayor, en el que se necesitaría cada vez más precisión para ir cada vez más lejos en la lista de preguntas. Sin embargo, si obtenemos algo nuevo de nuestro dispositivo en una etapa *finita* en precisión, al menos hasta que encontremos un algoritmo mejorado para la sucesión de preguntas; entonces tendríamos que ir a una mayor precisión para obtener algo que no pudiera decirnos nuestro algoritmo *mejorado.*

De todas formas, parecería aún que una precisión siempre creciente en un parámetro físico es una manera incómoda e insatisfactoria de codificar información. Mucho más preferible sería adquirir nuestra información en una forma *discreta* (o "digital"). Podrían conseguirse respuestas a preguntas cada vez más avanzadas en una lista examinando más unidades discretas cada vez, o quizá examinando un conjunto *fijo* de unidades discretas una y otra vez, en donde la ilimitada información requerida se extendería sobre intervalos de tiempo cada vez mayores. (Podemos imaginar estas unidades discretas constituidas por partes, cada una de ellas susceptible de un estado "sí" o "no", como los 0's y los 1's de las descripciones de las máquinas de Turing dadas en el capítulo II.) Para esto parece que necesitamos ciertos dispositivos que puedan adoptar (de manera distinguible) estados discretos y que, después de evolucionar de acuerdo con las leyes dinámicas, adoptarían de nuevo un estado de un conjunto de estados discretos. Si así fuera podríamos evitarnos examinar cada dispositivo con una precisión arbitrariamente alta.

Ahora bien, ¿realmente se comportan de esta forma los sistemas hamiltonianos? Sería necesario algún tipo de estabilidad del comportamiento, de modo que pudiera comprobarse claramente en cuál de estos estados discretos está nuestro dispositivo. Una vez que está en uno de estos estados queremos que siga en él (al menos durante un tiempo considerable) y no se deslice de uno de estos estados a otro. Además, si el sistema llega a estos estados con cierta imprecisión no es bueno que estas imprecisiones crezcan; antes bien, lo que realmente exigimos es que estas imprecisiones se *atenúen* con el tiempo. Ahora nuestro supuesto dispositivo tendría que estar constituido por partículas (u otras subunidades) que se deben describir en función de parámetros continuos, y cada estado "discreto" distinguible tendrá que cubrir cierto *rango* de estos parámetros continuos. (Por ejemplo, una posible manera de representar instancias discretas sería tener una partícula que pueda estar en una caja o en otra. Para especificar que la partícula está realmente en una de las cajas necesitamos decir que las coordenadas de posición de la partícula están dentro de ciertos límites.) Lo que esto significa, en términos del espacio de fases, es que cada una de nuestras opciones "discretas" debe corresponder a una *región* del espacio de fases, de modo que puntos diferentes del espacio de fases que están en la misma región corresponderán a la *misma* alternativa para nuestro dispositivo (fig. V.12). Supongamos ahora que el dispositivo comienza con su punto en el espacio de fases dentro de alguna región R_0 que corresponde a una de estas opciones. Consideramos que R_0 es arrastrada a lo largo del campo de vectores hamiltoniano a medida que transcurre el tiempo, hasta que, en el instante *t*, la región se ha convertido en R_t. Al representar esto imaginamos al mismo tiempo la evolución temporal de nuestro sistema para *todos* los posibles estados de partida que corresponden a esta misma alternativa. (Véase fig. V.13.)

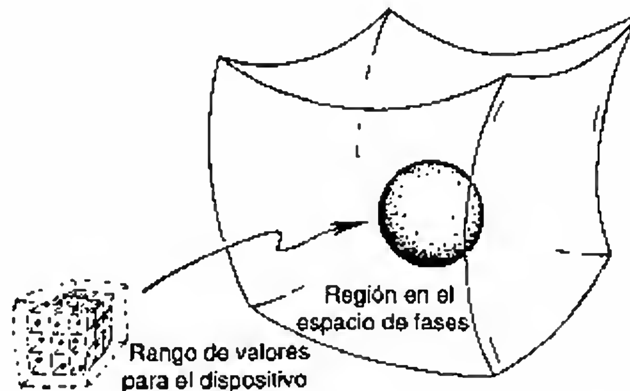


FIGURA V.12. *Una* región *en el espacio de fases corresponde a un intervalo de posibles valores de las posiciones y momentos de todas las paníc ulas. Una de estas regiones podría representar un estado distinguible (es decir, "alternativa") de algún dispositivo.*

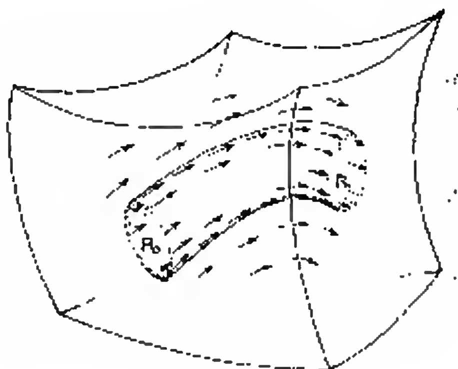


FIGURA V.13. *A medida que pasa el tiempo, una región de estados de fase* R_0 es *arrastrada a lo largo del campo de vectores hacia una nueva región* R_t. *Esto representará la evolución temporal de una alternativa particular para nuestro dispositivo.*

El problema de la *estabilidad* (en el sentido que aquí nos interesa) consiste en si la región *R,* permanece localizada, a medida que *t* crece, o si empieza a extenderse por el espacio de fases. Si estas regiones permanecen localizadas conforme avanza el tiempo, entonces tenemos una medida de estabilidad para nuestro sistema. Los puntos del espacio de fases que están estrechamente próximos (y por lo tanto corresponden a estados físicos detallados del sistema que se parecen estrechamente entre sí) permanecerán estrechamente próximos en el espacio de fases, y las imprecisiones en su especificación no se amplificarán con el tiempo. Cualquier dispersión indebida implicaría una impredecibilidad efectiva en el comportamiento del sistema.

¿Qué se puede decir sobre los sistemas hamiltonianos en general? ¿Tienden a dispersarse con el tiempo las regiones en el espacio de fases? Parecería que muy poco se puede decir sobre un problema de tal generalidad. Resulta, sin embargo, que existe un teorema muy bello, debido al famoso matemático francés Joseph Liouville (1809-1882), que nos dice que el *volumen* de cualquier región del espacio de fases debe permanecer constante en cualquier evolución hamiltoniana. (Por supuesto, toda vez que nuestro espacio de fases tiene una dimensión alta, éste tiene que ser un "volumen" en el sentido apropiado para una dimensión alta). Por consiguiente, el volumen de cada *R,* debe ser *igual* que el volumen de nuestra R_0 original. A primera vista esto parece responder afirmativamente a nuestra pregunta acerca de la estabilidad. En efecto, el tamaño —es decir, el volumen en el espacio de fases— de nuestra región no puede crecer, así que parece que nuestra región no puede dispersarse por el espacio de fases.

Sin embargo, esto es engañoso y reflexionando vemos que probablemente suceda el caso contrario. En la fig. V.14 he tratado de indicar la clase de comportamiento que uno esperaría en general. Podemos imaginar que la región inicial R_0 es una región de forma "razonablemente" pequeña, redondeada más que espigada, indicando que los estados que pertenecen a R_0 pueden definirse sin necesidad de recurrir a una precisión excesiva. Sin embargo, a medida que transcurre el tiempo, la región *R* comienza a distorsionarse y estirarse, siendo al principio quizá como una amiba, pero estirándose luego hasta grandes distancias en el espacio de fases y contorsionándose hacia adelante y hacia atrás en forma muy complicada. El volumen sigue siendo el mismo pero puede haberse dispersado sobre enormes regiones del espacio de fases.

Para una situación algo análoga piénsese en una pequeña gota de tinta colocada en un gran recipiente de agua. Aunque el volumen real del material que constituye la tinta permanece invariable, acabará por difundirse finalmente por todo el recipiente. En el caso del espacio de fases es probable que la región *R,* se comporte de una manera semejante. Puede que no se extienda por la *totalidad* del espacio de fases (que es la situación extrema que se conoce como "ergódica") pero es probable que se extienda sobre una región enormemente mayor que la de partida. (Véase también Davies, 1974.)

El problema es que la conservación del volumen no implica en absoluto conservación de *forma:* las regiones pequeñas tenderán a distorsionarse, y esta distorsión se magnifica en grandes distancias. El problema es mucho más grave en una dimensión alta que en una baja, puesto que hay muchas más "direcciones" en las que la región puede difundirse localmente. De hecho, lejos de ser una ayuda mantener la región R_t bajo control, el teorema de Liouville nos coloca realmente ante un problema fundamental.

Sin el teorema de Liouville podríamos suponer que esta indudable tendencia de una región a extenderse por el espacio de las fases podría quedar compensada, en circunstancias apropiadas, por una reducción del volumen global. Sin embargo, el teorema nos dice que esto es *imposible,* y tenemos que afrontar esta sorprendente consecuencia —que constituye una característica universal de todos los sistemas dinámicos (hamiltonianos) clásicos de tipo formal.[9]

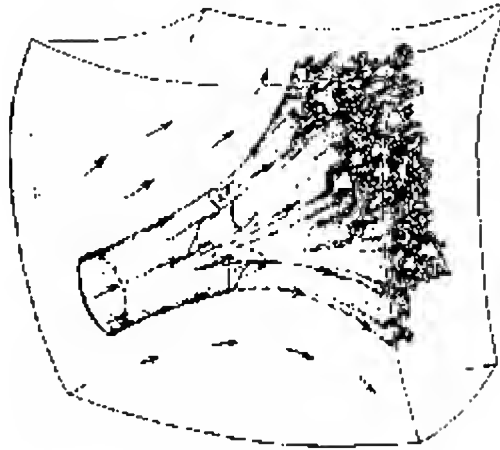


FIGURA V.14. *Pese al hecho de que el teorema de Liouville nos dice que el volumen del espacio de fases no cambia durante la evolución temporal, normalmente este volumen se dispersará* efectivamente *debido a la extrema complejidad de su evolución.*

Podemos preguntar, en vista de esta difusión por todo el espacio de fases, ¿cómo es posible hacer predicción alguna en mecánica clásica?

[9] La situación es, en realidad, "peor" en el sentido de que el volumen del espacio de fases de Liouville es sólo uno entre una familia entera de "volúmenes" de diferentes dimensiones (conocidos como invariantes de Poincaré) que permanecen constantes bajo evoluciones hamiltonianas. Sin embargo, he sido un poco injusto en la radicalidad de mis afirmaciones. Podemos imaginar un sistema en el que los grados de libertad físicos (que contribuyen a parte del volumen del espacio de fases) pueden estar "esparcidos" en alguna parte que no nos interesa (como radiación que escapa al infinito) de modo que el volumen en el espacio de fases en la parte que nos interesa puede reducirse.

Esta es, en verdad, una buena pregunta. Lo que esta difusión nos dice es que, independientemente de la precisión con que conozcamos el estado inicial de un sistema (dentro de límites razonables), las imprecisiones tenderán a crecer con el tiempo y nuestra información inicial puede hacerse casi inútil. En este sentido, la mecánica clásica es esencialmente *impredecible.* (Recuérdese el concepto de "caos" considerado antes.)

¿Cómo es entonces que la mecánica clásica ha resultado ser tan acertada? En el caso de la mecánica celeste (esto es, el movimiento de los cuerpos celestes bajo la acción de la gravedad) las razones parecen ser, en *primer lugar,* que nos interesamos en un número relativamente pequeño de cuerpos coherentes (el Sol, los planetas y los satélites) que presentan grandes diferencias en cuanto a su masa —de modo que de entrada podemos prescindir de los efectos perturbadores de los cuerpos de menor masa y tratar los mayores como si fueran sólo *unos pocos* cuerpos que actúan bajo la influencia de los demás— y en *segundo lugar,* que las leyes dinámicas que se aplican a las partículas individuales que constituyen estos cuerpos pueden verse operando también en el nivel de los propios cuerpos —de modo que, con bastante acierto, podemos considerar el Sol, los planetas y los satélites como partículas, sin tener que preocuparnos por todos los movimientos de cada una de las partículas que componen estos cuerpos celestes.[10] De nuevo salimos del paso considerando sólo unos "pocos" cuerpos, y la difusión en el espacio de fases no es importante.

Aparte de la mecánica celeste y del comportamiento de los proyectiles (que en realidad constituyen un caso particular de la mecánica celeste), y del estudio de los sistemas simples que abarcan un pequeño número de partículas, los principales modos de utilización de la mecánica newtoniana no parecen ser, ni mucho menos, de este tipo detallado "predictivo de forma determinista". Más bien, se utiliza el esquema general newtoniano para hacer modelos a partir de los cuales se pueden inferir propiedades globales de comportamiento. Algunas consecuencias precisas de las leyes, como la conservación de la energía, el momento y el momento angular tienen importancia en todas las escalas. Además, existen propiedades estadísticas que pueden combinarse con las leyes dinámicas que gobiernan las partículas individuales y pueden utilizarse para hacer predicciones globales acerca del comportamiento. (Véase lo concerniente a termodinámica en el capítulo VII; el efecto de difusión en el espacio de fases que ha ocupado nuestra atención se relaciona íntimamente con la segunda ley de la termodinámica, y con el cuidado debido estas ideas pueden usarse de formas verdaderamente predictivas.) El mismo cálculo notable de Newton para la velocidad del sonido en el aire (sutilmente corregido unos cien años más tarde por Laplace) fue un buen ejemplo de esto. Sin embargo, es muy raro que se utilice el determinismo inherente a la dinámica newtoniana (o, con más generalidad, hamiltoniana).

El efecto de difusión en el espacio de las fases tiene otra curiosa implicación. Nos dice, en efecto, que la *mecánica clásica no puede ser verdadera en nuestro mundo.* Exagero un tanto esta implicación, pero quizá no demasiado. La mecánica clásica puede dar cuenta perfectamente del

[10] Este segundo hecho, en particular, resulta de lo más afortunado para la ciencia, pues sin él el comportamiento dinámico de los cuerpos grandes podría haber sido incomprensible y hubiera dado pocas pistas sobre las leyes exactas que se aplicarían a las propias partículas. Me atrevo a conjeturar que la razón de que Newton haya insistido con tanta energía en su tercera ley era que, sin ella, esta transferencia de comportamiento dinámico desde los cuerpos microscópicos a los macroscópicos simplemente no se cumpliría. Otro hecho "milagroso", que fue vital para el desarrollo de la ciencia, es que la ley del inverso del cuadrado es la única ley de potencias (decreciente con la distancia) para la que las órbitas generales en torno a un cuerpo central son formas geométricas sencillas. ¿Qué hubiera hecho Kepler si la ley de fuerzas hubiera sido una ley del inverso de la distancia o del inverso del cubo?

comportamiento *de* los cuerpos fluidos —en particular de los gases, pero también de los líquidos en buena medida—, en donde sólo nos interesan propiedades globales "promedio" de los sistemas de partículas, pero tiene problemas para dar cuenta de la estructura de los sólidos, en los que es necesaria una organización más estructurada y detallada. Existe un problema en cómo puede mantener su forma un sólido cuando está compuesto de miríadas de partículas puntuales cuya organización se reduce continuamente debido a la difusión en el espacio de fases. Como ahora sabemos, se necesita la teoría cuántica para poder comprender adecuadamente la estructura real de los sólidos. Los efectos cuánticos pueden impedir de algún modo esta difusión en el espacio de fases. Este es un tema importante al que volveré más tarde (véanse los capítulos VIII y IX). Este asunto tiene también importancia para la cuestión de la construcción de una máquina computadora. La difusión en el espacio de fases es algo que debemos controlar. No se debe permitir que una región del espacio de fases que corresponde a un estado discreto de un dispositivo de computación (tal como la R_0 antes descritas) se disperse indebidamente. Recordemos que incluso el "computador de bolas de billar" de Fredkin-Toffoli requiere algunas *paredes sólidas* externas para que pueda funcionar. La "solidez" de un objeto compuesto de muchas partículas es algo que realmente necesita de la mecánica cuántica. Parece que incluso una máquina computadora "clásica" debe apropiarse de los efectos de la física cuántica si ha de funcionar eficazmente.

LA TEORÍA ELECTROMAGNÉTICA DE MAXWELL

Conforme a la imagen newtoniana del mundo, pensamos en minúsculas partículas que interaccionan gracias a las fuerzas que actúan a distancia —en donde las partículas, no enteramente puntuales, rebotan de vez en cuando unas en otras por contacto físico real. Como he establecido antes, las fuerzas de la electricidad y el magnetismo (cuya existencia había sido conocida desde la Antigüedad, y estudiadas con algún detalle por William Gilbert en 1600 y por Benjamín Franklin en 1752) actúan de un modo semejante a las fuerzas gravitatorias en que también decrecen de forma inversamente proporcional al cuadrado de la distancia, aunque son repulsivas en lugar de atractivas —es decir, actúan como si las partículas semejantes se repelieran— y es la carga eléctrica (y la intensidad del polo magnético), en lugar de la masa, la que mide la intensidad de la fuerza. En este nivel no hay dificultad en incorporar la electricidad y el magnetismo en el esquema newtoniano. También puede encajar aproximadamente el comportamiento de la luz (aunque con ciertas dificultades), ya sea considerando la luz como compuesta de partículas ("fotones", como se las llama ahora) o bien como un movimiento ondulatorio en algún medio, en cuyo caso debe considerarse este medio ("éter") como compuesto de partículas.

El hecho de que cargas eléctricas en movimiento puedan dar lugar a fuerzas magnéticas provocó alguna complicación adicional, pero no alteró el esquema general. Numerosos matemáticos y físicos (incluido a Gauss) habían propuesto sistemas de ecuaciones para los efectos de las cargas eléctricas en movimiento que habían parecido satisfactorias dentro del marco general newtoniano. El primer científico en desafiar seriamente la imagen newtoniana parece haber sido el gran teórico y experimental inglés Michael Faraday (1791-1867).

Para comprender la naturaleza de este desafío debemos entender primero el concepto de *campo*. Consideremos un campo magnético en primer lugar. Muchos lectores se habrán tropezado con el

comportamiento de las limaduras de hierro colocadas en una hoja de papel sobre un imán. Las limaduras se alinean de una manera sorprendente a lo largo de las llamadas "líneas de fuerza magnética". Imaginemos que las líneas de fuerza siguen presentes cuando las limaduras no están ahí. Constituyen lo que se llama un *campo magnético*. En cada punto del espacio este "campo" está orientado en cierta dirección, a saber, la dirección de la línea de fuerza que pasa por dicho punto. En realidad, lo que tenemos es un *vector* en cada punto de modo que el campo magnético nos proporciona un ejemplo de campo vectorial. (Podemos comparar esto con el campo de vectores hamiltoniano que consideramos en la sección anterior, pero ahora se trata de un vector en el espacio ordinario en lugar del espacio de fases.) Análogamente, un cuerpo eléctrico cargado estará rodeado por un tipo de campo diferente, conocido como un *campo eléctrico,* y un *campo gravitatorio* rodea del mismo modo cualquier cuerpo con masa. También estos son campos de vectores en el espacio.

Tales ideas eran conocidas mucho tiempo antes de Faraday y habían llegado a ser una buena parte del arsenal de los teóricos en mecánica newtoniana. Pero el punto de vista prevaleciente no consideraba que tales "campos" constituyeran en sí mismos una sustancia física real. Más bien debería pensarse que proporcionaban el necesario "balance de cuentas" para las fuerzas que actuarían si se colocara una partícula adecuada en varios puntos diferentes. Sin embargo, los profundos hallazgos experimentales de Faraday (con conductores en movimiento, imanes y similares) le llevaron a creer que los campos eléctricos y magnéticos son "sustancia" física *real* y, además, que campos eléctricos y magnéticos variables podrían a veces ser capaces de "empujarse entre sí" a través del espacio vacío, produciendo un tipo de onda incorpórea. Conjeturó que la propia luz podría consistir en dichas ondas. Semejante punto de vista estaría en desacuerdo con el saber newtoniano prevaleciente, en el que dichos campos no se consideraban "reales" en ningún sentido sino simplemente cómodos auxiliares matemáticos para la "verdadera" imagen newtoniana de la "realidad", vista como acción a distancia entre partículas puntuales.

Confrontado con los hallazgos experimentales de Faraday, y con algunos anteriores del notable físico francés André Marie Ampere (1775-1836) y otros, el gran físico y matemático escocés James Clerk Maxwell (1831-1879) se interrogó sobre la forma matemática de las ecuaciones para los campos eléctrico y magnético que surgían de aquellos hallazgos. Con un notable golpe de intuición propuso un cambio en las ecuaciones —quizá pequeño en apariencia, pero fundamental en sus consecuencias—. Este cambio no lo sugerían en absoluto los hechos experimentales conocidos (aunque era congruente con ellos); fue resultado de las propias exigencias teóricas de Maxwell, en parte físicas, en parte matemáticas y en parte estéticas. Una consecuencia de las ecuaciones de Maxwell era que los campos eléctrico y magnético se "empujan" a través del espacio vacío. Un campo magnético oscilante daría lugar a un campo eléctrico oscilante (lo que estaba implícito en los descubrimientos experimentales de Faraday) y este campo eléctrico oscilante daría lugar, a su vez, a un campo magnético oscilante (por una inferencia teórica de Maxwell), el que a su vez daría lugar a un campo eléctrico y así sucesivamente. (Véanse en las figs. VI.26, VI.27 las representaciones de estas ondas.) Maxwell pudo calcular la velocidad con la que se propagaría este efecto y encontró que sería la velocidad de la luz. Estas llamadas ondas *electromagnéticas* mostrarían también la propiedad de interferencia y la enigmática propiedad de polarización de la luz, conocidas desde hacia mucho tiempo (a ellas volveremos en el capítulo VI). Además de explicar las propiedades de la luz visible, para la que las ondas tendrían un intervalo concreto de longitudes de onda ($4 - 7 \times 10^{-7}$

m), se predecía la existencia de ondas electromagnéticas de otras longitudes de onda que se podrían producir mediante corrientes eléctricas en hilos conductores. En 1888, el notable físico alemán Heinrich Hertz estableció experimentalmente la existencia de tales ondas. La inspirada esperanza de Faraday había provisto de una base firme a las maravillosas ecuaciones de Maxwell.

Aunque no es necesario que apreciemos los detalles de las ecuaciones de Maxwell no estará de más echarles una ojeada:

$$\frac{1}{c^2}\cdot\frac{\partial E}{\partial t} = rotB - 4\pi j \qquad \frac{\partial B}{\partial t} = -rotE\,,$$

$$divE = 4\pi r \qquad divB = 0$$

Aquí, *E, B* y *j* son campos de vectores que describen el campo eléctrico, el campo magnético y la corriente eléctrica, respectivamente; r describe la densidad de carga eléctrica, y *c* es simplemente una constante: la velocidad de la luz.[11] No hay que preocuparse por los términos "rot" y "div" que simplemente se refieren a diferentes tipos de variación espacial. (Son ciertas combinaciones de los operadores de derivación parcial respecto a las coordenadas espaciales. Recuérdese la operación "derivada parcial", simbolizada por ∂, que ya encontramos a propósito de las ecuaciones de Hamilton.) Los operadores $\partial/\partial t$ que aparecen en el primer miembro de las dos primeras ecuaciones son, de hecho, iguales que el "punto" que utilizábamos en las ecuaciones de Hamilton, la diferencia es sólo técnica. Así, $\partial E/\partial t$ significa "el ritmo de variación del campo eléctrico" y $\partial B/\partial t$ significa "el ritmo de variación del campo magnético". La primera ecuación[*] nos dice cómo varía el campo eléctrico con el tiempo, en función de lo que hacen el campo magnético y la corriente eléctrica en ese momento; mientras que la segunda ecuación dice cómo varía el campo magnético con el tiempo en función de lo que hace el campo eléctrico en ese momento. La tercera ecuación es, hablando en términos generales, una forma codificada de la ley de la inversa del cuadrado que nos dice cómo debe relacionarse (en cada instante) el campo eléctrico con la distribución de cargas; mientras que la cuarta ecuación expresa lo que sería equivalente para el campo eléctrico, salvo que en este caso no existen "cargas magnéticas" (partículas aisladas "polo Norte" o "polo Sur").

Estas ecuaciones son semejantes a las de Hamilton en cuanto que dicen cuál debe ser el ritmo de variación con el tiempo de las cantidades pertinentes (aquí los campos eléctrico y magnético) en función de cuáles son sus valores en cualquier instante. Por lo tanto, las ecuaciones de Maxwell son *deterministas,* al igual que las teorías hamiltonianas ordinarias. La única diferencia —y *es* una diferencia importante— es que las ecuaciones de Maxwell son ecuaciones de *campo* y no ecuaciones para partículas, lo que significa que se necesita un número *infinito* de parámetros para describir el estado del sistema (los vectores del campo en cada punto del espacio), en lugar del número finito que se necesita en una teoría de partículas (tres coordenadas de posición y tres de momento para cada partícula). Por consiguiente, el espacio de fases para la teoría de Maxwell es

[11] Elegí las unidades de los diversos campos de manera que coincidan lo más posible con la forma en la que Maxwell presentó originalmente sus ecuaciones (excepto que su densidad de carga sería mi c^{-2}r). Para otras unidades los factores de *c* estarían distribuidos de otra manera.

[*] Fue la introducción de $\partial E/\partial t$ en esta ecuación lo que constituyó el golpe maestro de inferencia teórica de Maxwell. Todos los demás términos de las ecuaciones ya eran conocidos, de hecho, por evidencia experimental directa. El coeficiente $1/c^2$ es muy pequeño y por esta razón no se había observado experimentalmente aquel término. (Para conservar sentido de esta nota se ha respetado la forma de las ecuaciones de Maxwell que figuran en el original. Hay que señalar, no obstante, que en el sistema de unidades MKS tanto la derivada temporal de *E* como la de *B* van multiplicadas por el factor 1/c [N del T.]).

un espacio con un número *infinito* de dimensiones. (Como lo expresé antes, las ecuaciones de Maxwell pueden ser englobadas en el marco general hamiltoniano, pero este marco debe ser ligeramente ampliado debido a esta dimensionalidad infinita.)[12]

Además de lo que ya se ha visto, el *nuevo* ingrediente fundamental para nuestra imagen de la realidad física que presenta la teoría de Maxwell consiste en que ahora los *campos* deben tomarse en serio por propio derecho y no pueden considerarse como simples apéndices matemáticos a las partículas "reales" de la teoría newtoniana. En realidad, Maxwell demostró que cuando los campos se propagan como ondas electromagnéticas transportan con ellos cantidades definidas de *energía.* Él pudo proporcionar una expresión explícita para esta energía. El hecho notable de que la energía pueda transportarse realmente de un lugar a otro por estas ondas electromagnéticas "incorpóreas" fue, en efecto, confirmado experimentalmente mediante la detección de tales ondas por Hertz. Ahora nos es familiar —aunque no deje de resultar bastante notable— que las ondas de radio transporten realmente energía.

COMPUTABILIDAD Y ECUACIÓN DE ONDA

Maxwell dedujo de sus ecuaciones que en las regiones del espacio en donde no hay cargas o corrientes (es decir, si $j = 0$, $r = 0$ en las ecuaciones anteriores), todas las componentes de los campos eléctrico y magnético deben satisfacer una ecuación conocida como la *ecuación de onda.*[#] La ecuación de onda puede considerarse una "versión simplificada" de las ecuaciones de Maxwell puesto que es una ecuación para una *sola cantidad* en lugar de las seis componentes de los campos eléctrico y magnético. Su solución ejemplifica el comportamiento ondulatorio sin mayores complicaciones, tales como la "polarización" de la teoría de Maxwell (dirección del vector del campo eléctrico).

La ecuación de onda tiene tanto mayor interés para nosotros cuanto ha sido estudiada explícitamente en relación con sus propiedades de *computabilidad.* En efecto, Marian Boykan Pour-El e Ian Richards (1979 1981, 1982, cfr. también 1989) han podido demostrar que incluso si las soluciones de la ecuación de onda se comportan de modo *determinista* en el sentido ordinario —es decir, los datos proporcionados para un instante inicial dado determinarán la solución para cualquier otro instante— existen datos iniciales *computables,* de cierto tipo "peculiar", con la propiedad de que, para un tiempo posterior computable, el valor determinado del campo resulta ser realmente *no computable.* Por ello, las ecuaciones de una teoría de campos plausible (aunque no exactamente la teoría de Maxwell que realmente es válida en nuestro mundo) puede dar lugar, en el sentido de Pour-El y Richards, a una evolución no computable.

A primera vista este es un resultado bastante sorprendente —y parece contradecir mis conjeturas de la última sección acerca de la probable computabilidad de los sistemas hamiltonianos "razonables"—. Sin embargo, aunque el resultado de Pour-El y Richards es ciertamente sorprendente y de importancia matemática, no contradice dicha conjetura de una manera que

[12] En efecto, tenemos un número *infinito* de x_i,s y p_i,s; pero existe también la complicación de que no podemos utilizar simplemente los valores de los campos en esas coordenadas, siendo necesario un cierto "potencial" en el campo de Maxwell para que pueda aplicarse el esquema hamiltoniano.

[#] La ecuación de onda (o ecuación de D'Alembert) puede escribirse

$$\left\{\left(\frac{1}{c^2}\right)\left(\frac{\partial}{\partial t}\right)^2-\left(\frac{\partial}{\partial x}\right)^2-\left(\frac{\partial}{\partial y}\right)^2-\left(\frac{\partial}{\partial z}\right)^2\left(\frac{1}{c^2}\right)\right\}\varphi$$

tenga sentido físico aceptable. La razón es que su tipo "peculiar" de datos iniciales no "cambia suavemente"[13] como lo requiere normalmente un campo físico razonable. En realidad, Pour-El y Richards demuestran que la no computabilidad *no puede aparecer* para la ecuación de ondas si rechazamos este tipo de campos. En cualquier caso, incluso si se permitieran campos de este tipo, sería difícil ver cómo un "dispositivo" físico (¿quizá un cerebro humano?) pudiera hacer uso de tal "no computabilidad". Solamente podría venir a cuento cuando se admitieran medidas de una precisión arbitrariamente alta, lo que, como describí antes, no es físicamente muy realista. De todas formas, el resultado de Pour-El y Richards supone un comienzo intrigante para una importante área de investigación en la que se ha trabajado poco hasta ahora.

LA ECUACIÓN DE LORENTZ: LAS PARTICULAS DESBOCADAS

Las ecuaciones de Maxwell, tal como están, no son un sistema de ecuaciones completo. Nos proporcionan una maravillosa descripción del modo en que se propagan los campos eléctrico y magnético una vez *dadas* las distribuciones de cargas y corrientes eléctricas. Estas cargas nos vienen dadas físicamente en forma de *partículas cargadas* —principalmente electrones y protones, como ahora sabemos— y las corrientes se deben al movimiento de dichas partículas. Si sabemos dónde están estas partículas y cómo se están moviendo, entonces las ecuaciones de Maxwell nos dicen cómo se comportará el campo electromagnético. Lo que *no* nos dicen las ecuaciones de Maxwell es cómo se comportarán las propias partículas. En los días de Maxwell se conocía una respuesta parcial a esta pregunta, pero no se había establecido un sistema satisfactorio de ecuaciones hasta que, en 1895, el famoso físico holandés Hendrick Antoon Lorentz utilizó ideas próximas a las de la teoría de la relatividad especial para derivar las que ahora se conocen como *ecuaciones de movimiento de Lorentz* para una partícula cargada *(cfr.* Whittaker, 1910, pp. 310, 395). Estas ecuaciones nos dicen cómo varía continuamente la velocidad de una partícula cargada debido a los campos eléctrico y magnético en el punto en que se localiza la partícula.[14] Cuando se añaden las ecuaciones de Lorentz a las de Maxwell se obtienen las reglas de evolución temporal tanto de las partículas cargadas como del campo electromagnético.

Sin embargo, no todo marcha bien con este sistema de ecuaciones. Nos proporcionan excelentes resultados si los campos son muy uniformes hasta la escala del tamaño de las propias partículas (tomando este tamaño como el "radio típico" del electrón, alrededor de 10^{-15} m), y siempre que los movimientos de las partículas no sean demasiado violentos. Sin embargo, existe aquí una dificultad de *principio* que puede ser importante en otras circunstancias. Lo que las ecuaciones de Lorentz nos dicen que debemos hacer es examinar el campo electromagnético en el *punto* exacto en que se encuentra la partícula (y, en efecto, nos proporcionan una "fuerza" en dicho punto). ¿Qué punto debe tomarse si la partícula tiene un tamaño finito? ¿Debemos tomar el "centro" de la partícula, o bien debemos promediar el campo (de la "fuerza") sobre todos los puntos de la superficie? Esto podría suponer una diferencia si el campo *no* es uniforme a la

[13] Es decir, no dos veces diferenciable.

[14] Las ecuaciones de Lorentz nos dicen cuál es la *fuerza* sobre una partícula cargada, debida al campo electromagnético en el que está inmersa; entonces, si conocemos su masa, la segunda ley de Newton nos dice cuál es la aceleración de la partícula. Sin embargo, las partículas cargadas suelen moverse con frecuencia a velocidades próximas a la de la luz y los efectos de la relatividad especial comienzan a ser importantes, lo que afecta a cuál debe ser la masa de la partícula que debemos tomar (véase la próxima sección). Fueron razones de este tipo las que retrasaron el descubrimiento de la ley de las fuerzas que actúan sobre una partícula cargada hasta el nacimiento de la relatividad especial.

escala de la partícula. Hay otro problema más serio: ¿cuál *es* realmente el campo en la superficie de la partícula (o en el centro)? Recuérdese que consideramos una partícula *cargada.* Existirá un campo electromagnético *debido a la propia partícula*

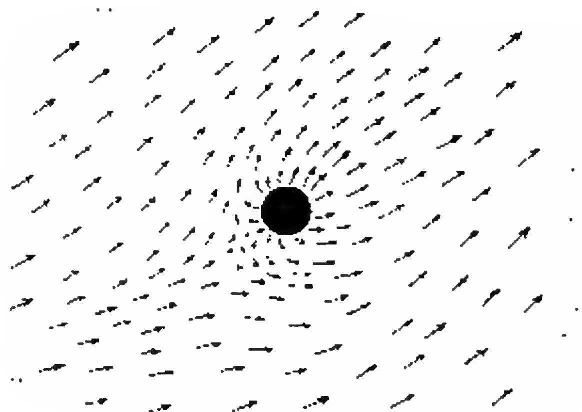


FIGURA V.15. *¿Cómo se aplican las ecuaciones de movimiento de Lorentz? Las fuerzas sobre una partícula cargada no pueden obtenerse simplemente examinando el campo "en" el sitio en el que se encuentra la partícula, ya que allí es dominante el propio campo de la partícula.*

que debe añadirse al "campo de fondo" en el que se sitúa la partícula. El campo de la propia partícula se hace enormemente intenso muy cerca de su "superficie" y supera con mucho a todos los demás campos de su vecindad. Además, el campo de la partícula apuntará más o menos directamente hacia fuera (o hacia dentro) en cada punto, de modo que el campo resultante real, al que se supone que responde la partícula, no será ni mucho menos uniforme sino que apuntará en direcciones diferentes en distintos lugares de la "superficie" de la partícula, y ya no digamos en su "interior" (fig. V.15). Ahora tiene que empezar a preocuparnos si las diferentes fuerzas sobre la partícula tenderán a girarla o a distorsionarla, y debemos preguntarnos qué propiedades elásticas tiene, etcétera, (surgen aquí cuestiones especialmente problemáticas en relación con la *relatividad,* con las cuales no abrumaré al lector). Evidentemente el problema es mucho más complicado de lo que parecía.

Quizá sea mejor considerar la partícula como de índole *puntual.* Pero esto lleva a otro tipo de problemas pues entonces el propio campo eléctrico de la partícula se hace *infinito* en su inmediata vecindad. Si, según las ecuaciones de Lorentz, la partícula tiene que responder al campo electromagnético en el que está inmersa, entonces, al parecer tendría que responder a un campo infinito. Para que la ley de la fuerza de Lorentz tenga sentido se necesita una manera de *restar el propio campo de la partícula* de modo que sólo quede un campo de fondo finito al que la partícula puede responder de modo inequívoco. El problema de cómo hacer esto fue resuelto en 1938 por Dirac (de quien volveremos a ocuparnos más adelante). Sin embargo, su solución llevó a Dirac a algunas conclusiones alarmantes. Encontró que para que el comportamiento de las partículas y campos esté determinado por sus datos iniciales es necesario conocer no sólo la posición y velocidad inicial de cada partícula sino también su *aceleración* inicial (situación un tanto anómala en el contexto de las teorías dinámicas corrientes). Para la mayoría de los valores de esta aceleración inicial la partícula se comporta finalmente de un modo totalmente disparatado, acelerándose espontáneamente hasta una velocidad que se aproxima muy

rápidamente a la de la luz. Estas son las "soluciones desbocadas" de Dirac y no corresponden a nada de lo que realmente sucede en la naturaleza. Debemos encontrar un modo de descartar las soluciones desbocadas escogiendo las aceleraciones iniciales del modo correcto. Esto puede hacerse siempre, pero sólo si aplicamos "preciencia"; es decir, debemos especificar las aceleraciones iniciales de una manera que prevea qué soluciones se harán finalmente soluciones desbocadas, para así evitarlas No es así ni mucho menos como se especifican los datos iniciales en un problema físico determinista estándar. Con el determinismo convencional aquellos datos pueden darse arbitrariamente, libres de cualquier requisito sobre cómo vaya a ser el comportamiento futuro. Aquí no sólo está el futuro completamente determinado por los datos que pueden especificarse en un instante del pasado, sino que la propia especificación de dichos datos está limitada de forma muy precisa por el requisito de que el comportamiento futuro sea "razonable".

Hasta aquí podemos llegar con las ecuaciones fundamentales clásicas. El lector se habrá dado cuenta de que el tema del determinismo y la computabilidad en las leyes de la física clásica se ha hecho inquietantemente confuso. ¿Tenemos realmente un elemento *ideológico* en las leyes físicas, en donde el futuro influye de algún modo sobre lo que está permitido que suceda en el pasado? Ciertamente los físicos no aceptan normalmente estas implicaciones de la *electrodinámica clásica* (la teoría de las partículas cargadas clásicas, y los campos eléctrico y magnético) como descripciones serias de la realidad. Su respuesta normal a las dificultades anteriores consiste en decir que con partículas cargadas individuales se está propiamente en el dominio de la *electrodinámica cuántica,* y no se puede confiar en obtener respuestas razonables utilizando un procedimiento estrictamente clásico. Esto es indudablemente cierto pero, como veremos más adelante, la *propia* teoría cuántica tiene problemas a este respecto. De hecho, Dirac había considerado el problema clásico de la dinámica de una partícula cargada precisamente *porque* pensaba que podría aportar intuiciones para resolver las dificultades fundamentales aun mayores del (más apropiado físicamente) problema cuántico. Más adelante habremos de enfrentar a los problemas de la teoría *cuántica*

LA RELATIVIDAD ESPECIAL DE EINSTEIN Y POINCARÉ.

Recordemos el principio de relatividad galileana que nos dice que las leyes físicas de Galileo y Newton permanecen totalmente invariantes si pasamos de un sistema de referencia en reposo a un sistema en movimiento. Esto implica que no podemos distinguir, mediante un simple examen del comportamiento dinámico de los objetos próximos a nosotros, si estamos en reposo o si nos movemos con velocidad uniforme en alguna dirección. (Recordemos el ejemplo del barco en el mar, que da Galileo). Pero supongamos que añadimos a estas leyes, las de Maxwell: ¿sigue siendo cierta la relatividad galileana? Recuérdese que las ondas electromagnéticas de Maxwell se propagan a una velocidad fija *c,* la velocidad de la luz. El sentido común parece decirnos que si estuviéramos viajando muy rápidamente en una dirección, entonces la velocidad de la luz en dicha dirección debería parecemos *reducida* por debajo de *c* (debido a que nos movemos para "alcanzar a la luz" en dicha dirección) y la velocidad aparente de la luz en la dirección contraria debería, en consecuencia, estar *incrementada* por encima de *c* (debido a que nos alejamos de la luz) —que es diferente del valor *fijo c* de la teoría de Maxwell—. De hecho, el sentido común tendría razón: las ecuaciones de Newton y Maxwell combinadas *no* satisfacen la relatividad galileana.

Fue la preocupación por estos temas la que llevó a Einstein, en 1905 —como, en efecto había llevado antes a Poincaré (en 1898-1905)—, a la teoría de la relatividad especial. Poincaré y Einstein descubrieron independientemente que las ecuaciones de Maxwell *también* satisfacen un Principio de relatividad *(cfr.* Pais, 1982); es decir, las ecuaciones tienen una propiedad análoga a la de permanecer invariantes si pasamos de un sistema de referencia en reposo a uno en movimiento, aunque las reglas Para esto son *incompatibles* con las de la física de Galileo y Newton. Para hacerlas compatibles sería necesario modificar uno u otro conjunto de ecuaciones o renunciar al principio de relatividad.

Einstein no tenía intención de renunciar al principio de relatividad. Su instinto físico le hacía insistir en que este principio debe ser válido Para las leyes físicas de nuestro mundo. Además, él sabía que, prácticamente en todos los fenómenos conocidos, la física de Galileo y Newton había sido verificada sólo para velocidades muy pequeñas comparadas con la velocidad de la luz, donde esta incompatibilidad no era significativa. El único fenómeno conocido con velocidades suficientemente grandes para que tales discrepancias fueran importantes era la *propia luz.* Sería, por lo tanto, el comportamiento de la luz el que nos informara sobre qué principio de relatividad debíamos adoptar, y las ecuaciones que gobiernan la luz son las de Maxwell. Por lo tanto, el principio que había que sustentar era el de relatividad de la teoría de Maxwell y, consiguientemente, había que modificar las leyes de Galileo y Newton.

Lorentz, antes que Poincaré y Einstein, también había respondido parcialmente estas cuestiones. Hacia 1895 Lorentz había aceptado la idea de que las fuerzas que ligan la materia eran de naturaleza electromagnética (como realmente resultaron ser) de modo que el comportamiento de los cuerpos materiales reales debería satisfacer leyes derivadas de las ecuaciones de Maxwell. Una consecuencia de ello resultaba ser que un cuerpo que se mueva con una velocidad comparable a la de la luz se contraerá ligeramente en la dirección del movimiento (la "contracción de Fitzgerald-Lorentz"). Lorentz había utilizado esto para explicar un enigmático descubrimiento experimental, el de Michelson y Morley en 1887, que parecía indicar que no pueden usarse los fenómenos electromagnéticos para determinar un sistema de referencia estático "absoluto". (Michelson y Morley demostraron que, contrariamente a lo que se esperaba, la velocidad aparente de la luz en la superficie de la Tierra no depende del movimiento de la Tierra alrededor del Sol.) ¿Siempre se comporta la materia de modo que no pueda detectarse localmente su movimiento (uniforme)? Esta fue la conclusión *aproximada* de Lorentz; además él se limitaba a una concreta teoría de la materia, en la que ninguna fuerza distinta de las electromagnéticas se consideraba significativa. Poincaré, siendo como era un matemático sobresaliente, pudo demostrar (en 1905) que la materia debe comportarse de una forma *precisa,* según el principio de relatividad inherente a las ecuaciones de Maxwell, para que localmente no pueda detectarse el movimiento uniforme. También obtuvo una buena comprensión de las implicaciones físicas de este principio (incluso la "relatividad de la simultaneidad" que consideraremos dentro de poco). Parece haberla considerado como sólo *una* posibilidad, y no compartía la convicción de Einstein de que algún principio de relatividad *deba* ser válido.

El principio de relatividad que satisfacen las ecuaciones de Maxwell —lo que ha llegado a conocerse como *relatividad especial*— es algo difícil de captar, y tiene muchas características no intuitivas que al principio son difíciles de aceptar como propiedades reales del mundo en que vivimos. De hecho, la relatividad especial no puede entenderse propiamente sin el ingrediente introducido en 1908 por el original e intuitivo geómetra ruso-germano Hermann Minkowski (1864-1909). Minkowski había sido uno de los profesores de Einstein en el Instituto Politécnico

de Zurich. Su nueva idea fundamental era que había que considerar el espacio y el tiempo en conjunto como una sola entidad: un *espacio-tiempo tetradimensional.* En 1908 Minkowski anunció en una famosa conferencia en la universidad de Gotinga:

> En lo sucesivo el espacio por sí mismo, y el tiempo por sí mismo, están condenados a desvanecerse en meras sombras, y sólo una especie de fusión entre los dos mantendrá una realidad independiente.

Tratemos de comprender las bases de la relatividad especial en función de este magnífico espacio-tiempo de Minkowski.

Una de las dificultades para comprender el concepto de espacio-tiempo es que tiene cuatro dimensiones, lo que lo hace difícil de imaginar. Sin embargo, después de haber sobrevivido a nuestro encuentro con el espacio de fases no tendremos ningún problema con tan sólo cuatro dimensiones. Como antes, haremos "trampa" y representaremos un espacio de dimensión menor, pero ahora el grado de la trampa es incomparablemente menos grave y, consecuentemente, nuestra representación será más aproximada. Dos dimensiones (una espacial y una temporal) bastarán para muchos propósitos, pero espero que el lector me permitirá ser algo más temerario y subir a tres (dos espaciales y una temporal). Esto proporcionará una representación muy buena y no será difícil aceptar que en principio las ideas se extiendan, sin muchos cambios, a la compleja situación tetradimensional. Lo que debemos tener en mente a propósito de un diagrama espacio-tiempo es que cada punto de la imagen representa un *suceso;* es decir, un punto en el espacio en un simple momento, un punto que tiene sólo una existencia *instantánea.* El diagrama completo representa toda la historia: pasado, presente y futuro. Una partícula, en tanto que persiste en el tiempo, viene representada no por un punto sino por una línea, llamada la *línea de universo.* Esta línea de universo —recta si la partícula se mueve uniformemente y curva si se acelera (es decir, si se mueve no uniformemente)— describe toda la historia de la existencia de la partícula.

En la fig. V.16 se representa un espacio-tiempo con dos dimensiones espaciales y una temporal. Imaginamos que hay una coordenada temporal estándar *t,* medida en la dirección vertical, y dos coordenadas espaciales *x/c* y *z/c,* medidas horizontalmente.* El cono en el centro es el cono de luz (futuro) del origen espacio-temporal *O.*

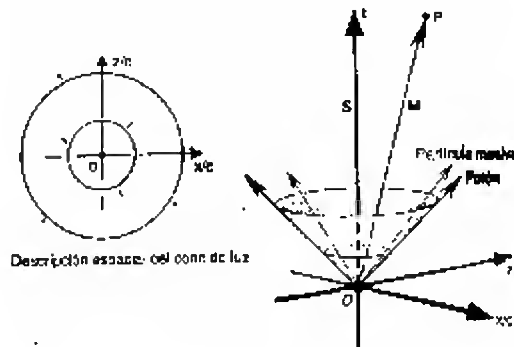


FIGURA V.16. *Cono de luz en el espacio-tiempo de Minkowski (con sólo dos dimensiones espaciales), que describe la historia del destello de una explosión que tiene lugar en el suceso* O, *origen del espacio-tiempo*

* La razón para dividir las coordenadas espaciales por *c* —la velocidad de la luz— es que las líneas de universo de los fotones tengan una inclinación conveniente: 45° respecto a la vertical; véase más adelante.

Para darnos cuenta de su significado imaginemos que una explosión tiene lugar en el suceso *O*. (Por lo tanto, la explosión ocurre en el origen espacial en el instante $t = 0$.) La historia de la luz emitida por la explosión es este cono de luz. En términos de dos dimensiones espaciales, la historia del destello de luz sería un círculo que se mueve hacia arriba con la velocidad fundamental de la luz *c*. En el espacio completo tridimensional sería una *esfera* que se expande a velocidad *c* —el frente de onda esférico de la luz— pero aquí estamos *suprimiendo* la dirección espacial y, de modo que sólo tenemos un círculo como las ondas circulares que emanan del punto de caída de una piedra en un estanque. Podemos ver este círculo en la imagen espacio-temporal si hacemos cortes horizontales sucesivos a través del cono que se mueve uniformemente hacia arriba. Estos planos horizontales representan distintas descripciones espaciales a medida que la coordenada temporal *t* crece. Ahora bien, una de las características de la teoría de la relatividad es que es imposible para una partícula material viajar más rápidamente que la luz (más adelante abundaremos sobre esto). Todas las partículas materiales provenientes de la explosión se retardan detrás de la luz. Esto significa, en términos espacio-temporales que las líneas de universo de todas las partículas emitidas en la explosión deben estar *dentro* del cono de luz.

Es a menudo más conveniente concebir la luz como *partículas*—llamadas fotones— que como ondas electromagnéticas. Por el momento podemos pensar en un fotón como un pequeño "paquete" de oscilaciones de alta frecuencia del campo electromagnético. El término es más apropiado físicamente en el contexto de las descripciones *cuánticas* que consideraremos en el próximo capítulo, pero los "clásicos" fotones también nos serán útiles aquí. En el espacio libre los fotones viajan siempre en línea recta con la velocidad fundamental *c*. Esto significa que en la imagen del espacio-tiempo de Minkowski la línea de universo de un fotón viene representada siempre como una línea recta inclinada a 45° de la vertical. Los fotones producidos en nuestra explosión en *O* describen el cono de luz centrado en *Ó*.

Estas propiedades deben ser válidas en general en todos los puntos del espacio-tiempo. No hay nada especial en el origen; el punto *O* no es diferente de ningún otro punto. Por lo tanto debe haber un cono de luz en cada punto del espacio-tiempo, con un significado idéntico al del cono de luz en el origen. La historia de cualquier destello de luz —o las líneas de universo de los fotones, si preferimos utilizar una descripción de la luz como partículas— está siempre sobre la superficie del cono de luz en cada punto, mientras que la historia de cualquier partícula material debe estar siempre en el interior del cono de luz en cada punto. Esto se ilustra en la fig. V.17. La familia de conos de luz en todos los puntos puede ser considerada como parte de la *geometría minkowskiana* del espacio-tiempo.

¿Qué es la geometría minkowskiana? La estructura de conos de luz es su aspecto más importante, pero la geometría minkowskiana no se reduce a esto.

Existe un concepto de "distancia" que tiene algunas analogías notables con la distancia en la geometría euclidiana.

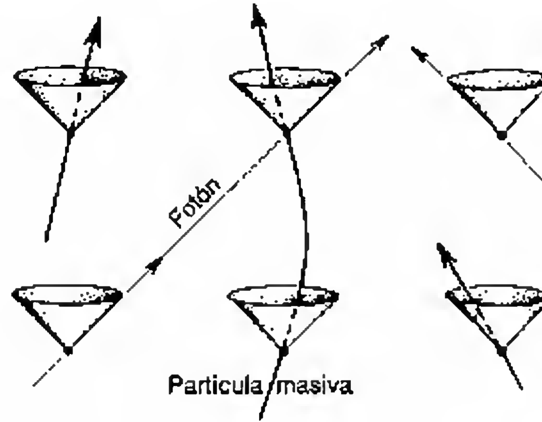


FIGURA V.17. *Imagen de la geometría minkowskiana.*

En la geometría euclidiana tridimensional, la distancia *r* de un punto al origen, en términos de las coordenadas cartesianas, está dada por

$$r^2 = x^2 + y^2 + z^2.$$

(Véase fig. V.18a. Esto es precisamente el teorema de Pitágoras —siendo quizá más familiar el caso de dos dimensiones.) En nuestra geometría minkowskiana tridimensional la expresión es formalmente muy parecida (fig. V.18b), siendo la diferencia esencial que ahora tenemos dos signos *de menos:*

$$s^2 = t^2 - (x/c)^2 - (z/c)^2.$$

Más correctamente, por supuesto, tendremos la geometría minkowskiana tetradimensional, y la expresión para la "distancia" es

$$s^2 = t^2 - (x/c)^2 - (y/c)^2 - (z/c)^2.$$

¿Cuál es el significado físico de la "distancia" *s* en esta expresión? Supongamos que el punto en cuestión —esto es, el punto *P*, de coordenadas *[t, x/c, y/c, z/c]*, (o *[t, x/c, z/c]* en el caso tridimensional; véase fig. V.16)— está dentro del cono de luz (futuro) de *O*. Entonces el segmento de línea recta *OP* puede representar parte de la historia de alguna partícula material —por ejemplo una partícula específica emitida por nuestra explosión.

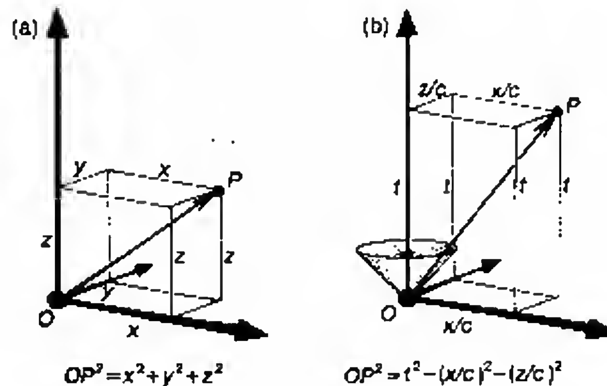


FIGURA V.18. *Comparación entre la "distancia" medida en (a) geometría euclidiana y (b) geometría minkowskiana (donde "distancia" significa "tiempo experimentado")*

La "longitud" minkowskiana *s* del segmento *OP* tiene una interpretación física directa: es el *intervalo temporal* experimentado por la partícula entre los sucesos O y P. Dicho de otro modo, si la partícula llevara consigo un poderoso y exacto reloj[15] entonces la diferencia entre tiempos que registraría en O y en P sería precisamente *s*. En contra de lo que se podría esperar, la magnitud de la coordenada *t no* describe "en sí misma el tiempo medido por un reloj preciso a menos que esté "en reposo" en nuestro sistema de coordenadas (es decir, con valores fijos *x/c, y/c, z/c* de las coordenadas), lo que quiere decir que el reloj tendría una línea de universo "vertical" en el diagrama. Por lo tanto, *"t"* significa "tiempo" sólo para observadores "en reposo" (esto es, con líneas de universo "verticales"). La medida *correcta* del tiempo para un observador en movimiento (que se aleja uniformemente del origen *O),* según la relatividad, viene dada por la cantidad *s.*

Esto es muy curioso y bastante distinto de la medida de tiempo galileo-newtoniana de "sentido común", que sería sencillamente el valor de la coordenada *t.* Nótese que la medida de tiempo relativista (minkowskiana) *s,* es siempre algo *menor* que t si no hay movimiento en absoluto (puesto que s^2 es menor que t^2, por la fórmula anterior, siempre que *x/c, y/c* y *z/c* no sean todas nulas). El movimiento (es decir, *OP* no está a lo largo del eje t) tenderá a "retardar" el reloj respecto a t —es decir, como se vería con respecto a nuestro sistema de coordenadas. Si la velocidad del movimiento es pequeña comparada con *c,* entonces s y t serán casi iguales, lo que explica por qué no somos conscientes directamente del hecho de que "los relojes en movimiento se atrasan". En el otro extremo, cuando la velocidad es la de la misma luz, *p* está entonces *sobre* el cono de luz; y encontramos s = 0. El cono de luz es precisamente el conjunto de puntos cuya "distancia" minkowskiana (es decir, "tiempo") a O es realmente cero. Así, un fotón no "experimentará" el paso del tiempo en absoluto. (No está permitido el caso aún *más* extremo, en el que *P* se mueve *fuera* del cono de luz, ya que ello conduciría a un s imaginario —la raíz cuadrada de un número negativo— y violaría la regla de que las partículas materiales o los fotones no pueden viajar más rápido que la luz.[*])

Esta noción de "distancia" minkowskiana se aplica igualmente a *cualquier* par de puntos del espacio-tiempo tal que uno está dentro del cono de luz del otro, de modo que una partícula puede viajar de uno a otro. Consideremos simplemente que *O* se desplaza a algún otro punto del espacio-tiempo. De nuevo, la distancia minkowskiana entre los puntos mide el intervalo de tiempo que experimenta un reloj que se mueve uniformemente de uno a otro. Cuando la partícula es un fotón, y la distancia minkowskiana se hace cero, tenemos dos puntos, uno de los cuales está *sobre* el cono de luz del otro. Este hecho sirve para *definir* el cono de luz de este punto.

La estructura básica de la geometría minkowskiana, con esta curiosa medida de "longitud" para las líneas de universo —interpretada como el *tiempo* medido (o "experimentado") por relojes físicos— contiene la misma esencia de la relatividad especial. Puede que el lector esté familiarizado, en particular, con lo que se conoce como "la paradoja de los gemelos" en relatividad: uno de los hermanos gemelos permanece en la Tierra mientras que el otro hace un viaje a una estrella cercana, yendo y volviendo a una gran velocidad que se aproxima a la de la

[15] De hecho, en cierto sentido, cualquier partícula mecano-cuántica en la naturaleza actúa por sí misma como tal reloj. Como veremos en el capítulo VI, existe una oscilación asociada a cada partícula cuántica, cuya frecuencia es proporcional a la masa de la partícula. Los relojes modernos de gran precisión (relojes atómicos, relojes nucleares) dependen en última instancia de este hecho.

[*] Sin embargo, para sucesos separados por valores negativos de s^2, la cantidad $c\sqrt{-(s)^2}$ tiene un significado, a saber, el de distancia *ordinaria* para aquel observador para quien los sucesos parecen simultáneos *(cfr.* más adelante).

luz. A su regreso, se descubre que los dos hermanos han envejecido de manera diferente: el viajero se encuentra aún joven mientras que el hermano que se quedó en casa es un viejo. Esto se describe fácilmente en términos de la geometría de Minkowski y se ve por qué, aunque es un fenómeno enigmático, no es realmente una paradoja. La línea de universo *AC* representa al gemelo que se queda en casa, mientras que el viajero tiene una línea de universo compuesta de dos segmentos *AB* y *BC* que representan las etapas de ida y vuelta del viaje (véase fig. V.19). El gemelo que se queda en casa experimenta un tiempo medido por la distancia minkowskiana *AC,* mientras que el viajero experimenta un tiempo dado por la suma[16] de las dos distancias de Minkowski *AB* y *BC.* Estos tiempos no son iguales sino que encontramos

$$AC > AB + BC,$$

lo que demuestra que realmente el tiempo experimentado por el que se queda en casa es mayor que el del viajero.

La desigualdad superior se parece bastante a la bien conocida *desigualdad del triángulo* de la geometría euclidiana ordinaria, a saber (siendo ahora *A, B* y *C* tres puntos en el espacio *euclidiano):*

$$AC < AB + BC,$$

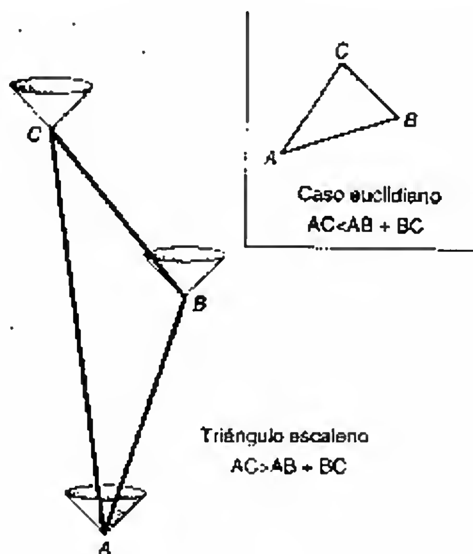


FIGURA V.19. *La llamada "paradoja de los gemelos " de la relatividad especial se comprende en términos de una desigualdad triangular minkowskiana. (Se da también el caso euclidiano para fines de comparación.)*

que afirma que la suma de dos lados de un triángulo es siempre *mayor* que el tercero. *Esto* no se considera una paradoja. Estamos perfectamente familiarizados con la idea de que la medida de la distancia euclidiana a lo largo de un camino de un punto a otro (aquí de A a C) depende del camino concreto que tomemos. (En este caso, los dos caminos son *AC* y la ruta mayor quebrada *ABC.*) Este ejemplo es un caso particular del hecho de que la distancia más corta entre dos

[16] Acaso el lector se preocupe por el hecho de que, puesto que en la línea de universo *de* viajero hay una "esquina" en *B,* el viajero sufre, como se muestra, una aceleración infinita en el suceso B. Esto es accesorio. Con una aceleración finita, la línea de universo del *viajero* tiene la esquina en *B* suavizada, y esto apenas supone diferencia en el tiempo total que experimenta, que aún se mide por la "longitud" minkowskiana de la línea de universo total.

puntos (aquí *A* y C) es la medida a lo largo de la línea recta que los une (la línea AC). El sentido contrario en el signo de la desigualdad en el caso *minkowskiano* aparece debido a los cambios de signo en la definición de "distancia", de modo que la *AC* minkowskiana es "más larga" que la ruta combinada *ABC.* Asimismo, esta "desigualdad triangular" minkowskiana es un caso particular de un resultado más general: la *más larga* (en el sentido de que se experimenta un tiempo mayor) de las líneas de universo entre dos puntos es la recta (es decir, no acelerada). Si dos gemelos empiezan en el mismo suceso *A* y terminan en el mismo suceso *C,* en donde el primer gemelo se mueve directamente de A a C sin aceleraciones pero el segundo acelera, entonces el primero experimentará siempre un intervalo de tiempo mayor cuando se encuentren de nuevo.

Puede parecer extravagante introducir un concepto tan extraño de medición del tiempo, en abierto contraste con nuestras nociones intuitivas. Sin embargo, hay ahora una enorme cantidad de datos experimentales en su favor. Por ejemplo, existen muchas partículas subatómicas que se desintegran (es decir, se deshacen en otras partículas) en una escala de tiempo definida. A veces tales partículas viajan a velocidades muy próximas a la de la luz (p. ej. los rayos cósmicos que llegan a la Tierra desde el espacio exterior, o las partículas aceleradas mediante aparatos construidos por el hombre), y sus tiempos de desintegración se alargan en la medida exacta que se deduce a partir de las consideraciones anteriores. Más impresionante aún es el hecho de que ahora pueden hacerse relojes (los "nucleares") tan precisos que estos efectos de retardo temporal son detectables *directamente* mediante relojes transportados en aviones rápidos en vuelo bajo, en concordancia con la medida de "distancia" minkowskiana *s,* y no con t. (Estrictamente hablando, si tomamos en cuenta la *altitud* del avión aparecerán pequeños efectos gravitatorios adicionales de la relatividad *general,* pero éstos concuerdan también con las observaciones; véase la próxima sección.) Además, existen muchos otros efectos íntimamente relacionados con el marco global de la relatividad especial, que continuamente reciben verificación detallada. Uno de éstos, la famosa relación de Einstein

$$E = mc^2,$$

que equipara energía y masa, tendrá para nosotros algunas implicaciones seductoras al final de este capítulo.

No he explicado aún de qué modo se incorpora el principio de relatividad a este esquema de cosas. ¿Cómo es que observadores que se mueven con velocidades uniformes diferentes pueden ser *equivalentes* para la geometría minkowskiana? ¿Cómo puede el eje temporal de la fig. V.16 ("observador en reposo") ser completamente equivalente a alguna otra línea de universo recta, digamos la recta *OP* extendida ("observador en movimiento")? Pensemos primero en la geometría *euclidiana.* Evidentemente dos líneas rectas cualesquiera son equivalentes con respecto a la geometría en general. Podemos imaginar que "deslizamos rígidamente sobre sí mismo" todo el espacio euclidiano hasta que una de las líneas rectas toma la posición de la otra. Pensemos en el caso bidimensional, el *plano* euclidiano. Podemos imaginar que movemos una hoja de papel rígidamente sobre una superficie plana de modo que una línea recta cualquiera dibujada en el papel llegue a coincidir con una línea recta dada en la superficie. Este movimiento rígido conserva la estructura de la geometría. Algo similar es válido para la geometría minkowskiana, aunque

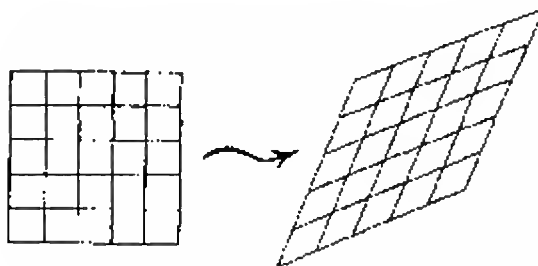


FIGURA V.20. *Movimiento de Poincaré en dos dimensiones espacio-temporales.*

esto es menos obvio y tenemos que tener cuidado en lo que entendemos por "rígido". Ahora, en lugar de en una hoja de papel deslizante debemos pensar en algún material de un tipo peculiar —tomando, para mayor sencillez, el caso *bidimensional*— en el que las líneas inclinadas 45° conservan esta inclinación mientras que el material puede estirarse en una dirección a 45° y contraerse en la otra dirección a 45°. Esto se ilustra en la fig. V.20. En la fig. V.21 he tratado de indicar lo que ocurre en el caso tridimensional. Este tipo de "movimiento rígido" del espacio de Minkowski —llamado *movimiento de Poincaré* (o movimiento no homogéneo de Lorentz)— puede parecer no muy "rígido" pero conserva todas las distancias minkowskianas, y es precisamente "conservación de las distancias" lo que significa la palabra "rígido" en el caso euclidiano. El principio de relatividad especial asegura que la física es invariante bajo tales movimientos de Poincaré del espacio-tiempo. En particular, el observador "en reposo" S, cuya línea de universo es el eje temporal de nuestra representación de Minkowski (fig. V.16), tiene una física completamente equivalente a la del observador "en movimiento" M con una línea de universo a lo largo de *OP.*

Cada plano de coordenadas t = constante representa el "espacio" en cada "instante" para el observador **S**, es decir, es una familia de sucesos que él consideraría *simultáneos* (esto es, que tienen lugar todos ellos en el "mismo instante"). Llamaremos a estos planos *espacios simultáneos* de **S**. Cuando pasamos a otro observador **M** debemos cambiar nuestra familia original de espacios simultáneos por una nueva familia mediante un movimiento de Poincaré que proporciona los espacios simultáneos Para **M.**[17] Nótese que los espacios simultáneos de **M** aparecen "inclinados" en la fig. V.21.

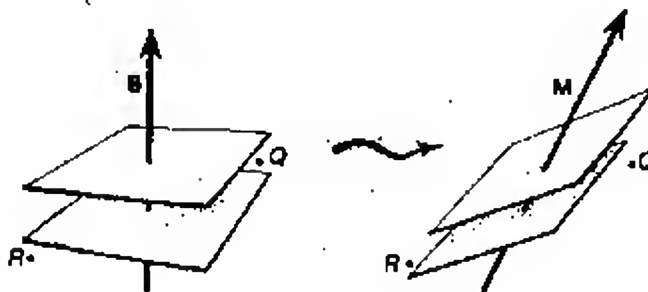


FIGURA V.21. *Movimiento de Poincaré en tres dimensiones espacio-temporales. El diagrama de la izquierda muestra espacios simultáneos para S y el de la derecha, espacios simultáneos para M. Nótese que S piensa que* R *precede a* Q, *mientras que M piensa que* Q *precede a* R. *(El*

[17] Estos son los espacios de sucesos que M juzgaría simultáneos según la *definición* de *simultaneidad de Einstein,* que utiliza las señales luminosas enviadas por M y devueltas a M desde los sucesos en cuestión. Véase, por ejemplo, Rindler (1982).

movimiento se considera aquí pasivo, es decir, que sólo afecta a las descripciones diferentes que los dos observadores S y M harían de un mismo espacio-tiempo.

Podría parecer que esta inclinación está en la dirección errónea si pensamos en función de movimientos rígidos en la geometría euclidiana, pero es la que debemos esperar en el caso minkowskiano. Mientras **S** piensa que todos los sucesos de cualquier plano t = constante ocurren simultáneamente, **M** tiene una visión diferente: para él son los sucesos de cada uno de sus espacios simultáneos "inclinados" los que parecen ser simultáneos. La geometría minkowskiana no contiene, en sí misma, un único concepto de "simultaneidad" sino que cada observador en movimiento uniforme lleva consigo su propia idea de lo que significa "simultáneo".

Consideremos los dos sucesos *R* y *Q* en la fig. V.21. Según *S*, el suceso *R* tiene lugar antes que el suceso *Q,* debido a que *R* está en un espacio simultáneo anterior al de *Q,* pero según **M** es al revés, *Q* está en un espacio simultáneo anterior al de *R*. Así, para un observador, el suceso *R* tiene lugar antes que *Q*, pero para el otro es *Q* el que tiene lugar antes que *R.* (Esto puede suceder sólo debido a que *R* y *Q* están *espacialmente separados,* lo que significa que cada uno de ellos está fuera del cono de luz del otro, de modo que ninguna partícula material ni ningún fotón puede viajar de un suceso a otro.) Incluso con velocidades relativas pequeñas pueden ocurrir diferencias significativas en el orden temporal para sucesos a grandes distancias. Imaginemos que dos personas que caminan lentamente se cruzan en la calle. Los sucesos en la galaxia Andrómeda (la más cercana de las grandes galaxias a nuestra propia Vía Láctea, a unos 20.000.000.000.000.000.000 kilómetros de distancia) que las dos personas juzgan simultáneos con el momento en que ellas se cruzan podrían tener una diferencia de varios días (fig. V.22).

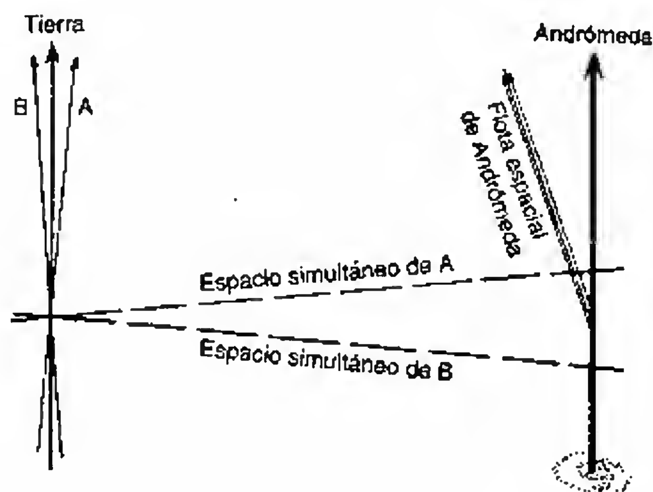


FIGURA V.22. *Dos personas A y B pasean lentamente, pero tienen diferentes visiones acerca de si ya ha partido una flota espacial de Andrómeda en el momento en que ellas se cruzan.*

Para una de las personas la flota espacial lanzada con la misión de acabar con la vida en el planeta Tierra está ya en camino; mientras que para la otra, la decisión de enviar esa flota todavía no se ha tomado.

LA RELATIVIDAD GENERAL DE EINSTEIN

Recuérdese que Galileo afirmó que todos los cuerpos caen con la misma velocidad en un campo gravitatorio. (Fue más una intuición, que una observación directa ya que, debido a la resistencia del aire, las plumas y las piedras *no* caen a la vez. La intuición de Galileo fue el darse cuenta de que si se pudiera anular la resistencia del aire, caerían a la vez.) Se necesitaron tres siglos para que la profunda significación de esta intuición fuese cabalmente entendida y constituyese la piedra angular de una gran teoría. Esta teoría fue la relatividad general de Einstein, una extraordinaria descripción de la gravitación que, como lo entenderemos en un momento, necesita el concepto de un *espacio-tiempo curvo* para su comprensión.

¿Qué tiene que ver la intuición de Galileo con la idea de "curvatura del espacio-tiempo"? ¿Cómo es posible que tal idea, en apariencia tan diferente del esquema de Newton en el que las partículas se aceleran bajo la acción de las fuerzas gravitatorias ordinarias, pudiera reproducir, *e* incluso mejorar, toda la soberbia precisión de dicha teoría? Además, ¿realmente puede ser cierto que la antigua intuición de Galileo contuviese algo que *no* fuera incorporado posteriormente a la teoría de Newton?

Permítaseme empezar por la última pregunta puesto que es la más fácil de responder. ¿Qué es, según la teoría de Newton, lo que gobierna la aceleración de un cuerpo sometido a la gravedad? En primer lugar tenemos la fuerza gravitatoria sobre dicho cuerpo, que la ley de atracción gravitatoria de Newton nos dice debe ser *proporcional a la masa del cuerpo.* En segundo lugar está la cantidad en que se acelera el cuerpo *dada* la fuerza que actúa sobre él, cantidad que, por la segunda ley de Newton, *es inversamente proporcional a la masa del cuerpo.* El hecho del que depende la intuición de Galileo es que la masa que aparece en la ley de la fuerza gravitatoria de Newton es la *misma* que la masa de la segunda ley de Newton. (Podría decirse "proporcional a" en lugar de "la misma que".) Esto es lo que asegura que la aceleración del cuerpo sometido a la gravedad es realmente *independiente* de su masa. No hay nada en el esquema general de Newton que exija que estos dos conceptos de masa sean el mismo. Newton sencillamente lo *postuló.* De hecho, las fuerzas eléctricas son semejantes a las fuerzas gravitatorias en que *ambas* son fuerzas del tipo de la inversa del cuadrado, pero la fuerza depende ahora de la *carga eléctrica* que es totalmente diferente de la *masa* de la segunda ley de Newton. La intuición de Galileo no se aplicaría a las fuerzas eléctricas: los objetos (esto es, los objetos cargados) "arrojados" en un campo eléctrico *no* "caen" todos con la misma velocidad.

Por el momento, *aceptemos* simplemente la intuición de Galileo —para el movimiento bajo la acción de la *gravedad*— y busquemos sus consecuencias. Imaginemos a Galileo arrojando dos piedras desde la torre inclinada de Pisa. Si hubiera habido una cámara de video en una de las piedras, apuntando hacia la otra, entonces la imagen que proporcionaría sería la de una piedra suspendida en el espacio, aparentemente *inafectada* por la gravedad (fig. V.23). Esto sucede precisamente debido a que todos los objetos sometidos a la gravedad caen con la misma velocidad.

Aquí pasamos por alto la resistencia del aire. Los vuelos espaciales nos ofrecen ahora una mejor verificación de estas ideas, puesto que no hay aire en el espacio exterior. Ahora, "caída" en el espacio significa sencillamente seguir la órbita apropiada bajo la gravedad. No es necesario que esta "caída" sea en línea recta hacia abajo, hacia el centro de la Tierra. Puede haber también una componente horizontal del movimiento. Si esta componente horizontal es suficientemente

grande, entonces se puede "caer" alrededor de la Tierra sin aproximarse al suelo. Viajar en una órbita libre bajo la acción de la gravedad es solamente una especial (y muy costosa) manera de "caer". Ahora, igual que en la imagen de la videocámara anterior, un astronauta durante un "paseo espacial" ve su nave como si estuviera suspendida delante de él, aparentemente inafectado por la fuerza gravitacional del enorme globo de la Tierra que se ve en el fondo. (Véase fig. V.24.) Por consiguiente, podemos eliminar localmente los efectos de la gravedad pasando al "sistema de referencia acelerado" en caída libre.

La gravedad puede así *cancelarse* mediante la caída libre debido a que los efectos del campo gravitatorio son precisamente iguales a los de una aceleración. De hecho, si usted está en el interior de un ascensor que acelera hacia arriba, simplemente experimentará un incremento del campo gravitatorio aparente; y si acelera hacia abajo, una disminución. Si se rompiera el cable que suspende el ascensor, entonces (haciendo a un lado la resistencia del aire y los efectos de rozamiento) la aceleración resultante hacia abajo neutralizaría completamente el efecto de la gravedad, y los ocupantes del ascensor parecerían flotar libremente —como el astronauta anterior— hasta que el ascensor golpeara el suelo. Incluso en un tren o en un avión las aceleraciones pueden ser tales que nuestras sensaciones sobre la intensidad y dirección de la gravedad pueden sugerirnos un "abajo" diferente del que nos indican nuestros sentidos. Esto

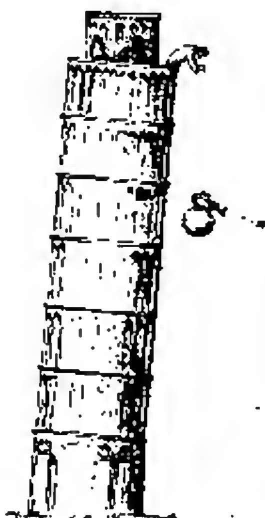


FIGURA V.23. *Galileo dejando caer dos Piedras (y una videocámara) desde la torre inclinada de Pisa.*



FIGURA V.24. *El astronauta ve su vehículo espacial suspendido ante él, aparentemente inafectado por la gravedad.*

es debido a que los efectos aceleracionales y gravitatorios *son* exactamente iguales, de modo que las sensaciones son incapaces de distinguir uno del otro. Este hecho —que los efectos locales de la gravitación son equivalentes a los de un sistema de referencia acelerado— es lo que Einstein denominó *principio de equivalencia.*

Las consideraciones anteriores son "locales". Sin embargo, si se nos permite hacer medidas (no completamente locales) de precisión suficiente, podemos, en principio, descubrir una *diferencia* entre un campo gravitatorio "verdadero" y una pura aceleración. En la fig. V.25 he mostrado, de

manera un poco exagerada, cómo una disposición de partículas que forman inicialmente una superficie esférica, cayendo libremente bajo la acción de la gravedad terrestre, empezará a ser afectada por la *no uniformidad* del campo gravitatorio (newtoniano). El campo es no uniforme en dos aspectos. En primer lugar, puesto que el centro de la Tierra está a una distancia finita, las partículas más próximas a la superficie de la Tierra se acelerarán hacia abajo más rápidamente que las que están más altas (recordemos la ley de la inversa del cuadrado de Newton). En segundo lugar, y por la misma razón, habrá ligeras diferencias en la *dirección* de esta aceleración para diferentes desplazamientos horizontales de las partículas. Debido a esta no uniformidad, la forma esférica comienza a distorsionarse ligeramente, transformándose en un elipsoide. Se alarga en dirección hacia el centro de la Tierra (y también en la dirección contraria), puesto que las partes más próximas al centro experimentan una aceleración ligeramente mayor que las partes más distantes; se estrecha en las direcciones horizontales, debido al hecho de que las aceleraciones dirigidas hacia el centro de la Tierra tienen una pequeña componente horizontal hacia adentro.

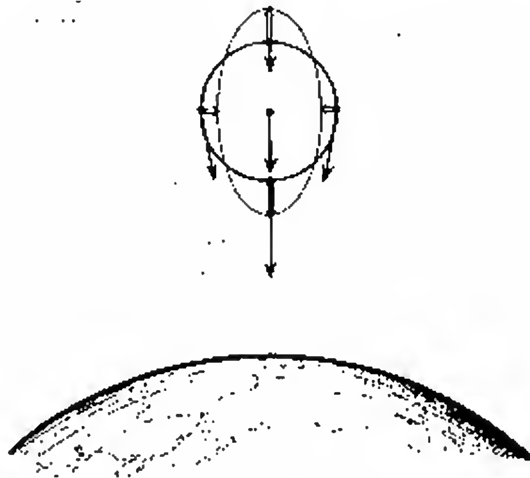


FIGURA V.25. *El efecto de marea. Las flechas dobles muestran la aceleración relativa.* (WEYL).

Este efecto distorsionante se conoce como el efecto *de marea* de la gravedad. Si reemplazamos el centro de la Tierra por la Luna, y la superficie esférica de partículas por la superficie de la Tierra, entonces tenemos precisamente la influencia de la Luna sobre las mareas en la Tierra, con abultamientos que se producen tanto en la dirección de la Luna como en la contraria. El efecto de marea es una característica general de los campos gravitatorios que no puede ser "eliminada" por la caída libre. Este efecto mide la no uniformidad del campo gravitatorio. (La *magnitud* de la distorsión de marea decrece según el inverso del *cubo* de la distancia al centro de atracción, en lugar de hacerlo según el inverso del cuadrado.)

La ley de la fuerza gravitatoria inversa del cuadrado de Newton resulta tener una interpretación simple en términos de este efecto de marea: el *volumen* del elipsoide en que se transforma la esfera inicial[18] es igual al de la esfera original, si se considera que la superficie esférica rodea a un vacío. Esta propiedad del volumen es característica de la ley del inverso del cuadrado; no se da para ninguna otra ley de fuerzas. Supongamos, a continuación, que la superficie esférica no rodea un vacío sino a una cierta masa *M*. Ahora habrá un componente adicional hacia adentro de

[18] Esta es la derivada *segunda* (o "aceleración") inicial de la forma con respecto al tiempo- El ritmo de variación (o "velocidad") de la forma se toma inicialmente como nulo, Puesto que la esfera está en reposo.

la aceleración, debido a la atracción gravitatoria de esta masa. El volumen del elipsoide en que se transforma la superficie esférica que forman las partículas inicialmente, se *contrae* — precisamente en una cantidad que es *proporcional* a *M.* Un ejemplo del efecto de reducción de volumen ocurriría si consideráramos esta superficie esférica rodeando la Tierra a una altura constante (fig. V.26). Entonces la aceleración hacia abajo (esto es, hacia adentro) ordinaria debida a la gravedad terrestre es la que causa la reducción del volumen de nuestra esfera. Esta propiedad de reducción de volumen codifica la parte restante de la ley de la fuerza gravitatoria de Newton, a saber, que la fuerza es proporcional a la masa del *cuerpo atrayente.*

Tratemos de obtener una imagen espacio-temporal de esta situación. En la fig. V.27 he indicado las líneas de universo de las partículas de nuestra superficie esférica (dibujada como un círculo en la fig. V.25), en donde he hecho la descripción en un sistema para el que el punto en el centro de la esfera parece estar en reposo ("caída libre"). El punto de vista de la relatividad general consiste en considerar los movimientos de caída libre como "movimientos naturales", los análogos del "movimiento uniforme en línea recta" que tenemos en la física sin gravedad.

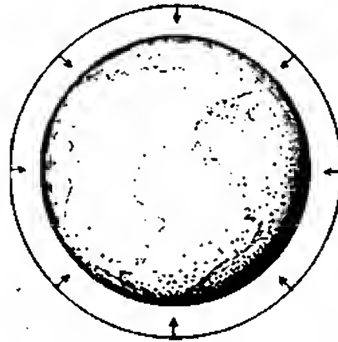


FIGURA V.26. *Cuando la Superficie esférica rodea materia (en este caso la Tierra) hay una aceleración neta hacia adentro.(RICCI)*

Así *tratamos* de considerar la caída libre como descrita por líneas de universo "rectas" en el espacio-tiempo. Sin embargo, por el aspecto de la fig. V.27, sería confuso utilizar *la palabra* "recta" para esto y, como cuestión de terminología, llamaremos geodésicas en el espacio-tiempo a las líneas de universo de las partículas en caída libre.

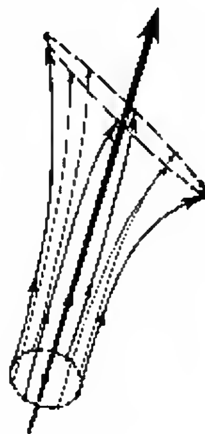


FIGURA V.27. *Curvatura espacio-temporal: el efecto de marea mostrado en el espacio-tiempo.*

¿Es ésta una buena terminología? ¿Qué entendemos normalmente por una "geodésica"? Examinemos una analogía para una superficie curva bidimensional. Las geodésicas son las curvas en dicha superficie que (localmente) son los caminos más cortos. Así, si pensamos en un trozo de cuerda estirado sobre la superficie (y no demasiado largo, pues podría escurrirse) entonces la cuerda coincidirá con una geodésica de la superficie. En la fig. V.28 he indicado dos ejemplos de superficies, la primera con lo que se llama "curvatura positiva" (como la superficie de una esfera) y la segunda con "curvatura negativa" (una superficie en forma de silla de montar). Para la superficie de curvatura positiva, dos geodésicas próximas que comienzan paralelas entre sí empezarán a curvarse *acercándose* si seguimos a lo largo de ellas; para curvatura negativa, empezarán a curvarse *apartándose* una de la otra. Si imaginamos que las líneas de universo de partículas en caída libre son en algún sentido análogas a las geodésicas en una superficie, entonces vemos que hay una estrecha analogía entre el efecto de marea gravitatorio, descrito arriba, y los efectos de curvatura de una superficie, pero ahora *ambos* efectos de curvatura, positiva y negativa, están presentes. Miremos las figs. V.25 y V.27. Vemos que nuestras "geodésicas" espacio-temporales empiezan a *separarse* en una dirección (cuando están alineadas en la dirección de la Tierra) —como sucede con la superficie de curvatura *negativa* de la fig. V.28— y comienzan a *acercarse* en otras direcciones (cuando se desplazan horizontalmente en relación con la Tierra), como sucede con la superficie de curvatura *positiva* de la fig. V.28. Así, nuestro espacio-tiempo parece poseer realmente una "curvatura", análoga a la de nuestras dos superficies pero más complicada debido a la mayor dimensionalidad, e intervienen mezclas de curvaturas tanto positivas como negativas para diferentes desplazamientos. Esto muestra cómo puede utilizarse un concepto de "curvatura" del espacio-tiempo para describir la acción de campos gravitatorios.

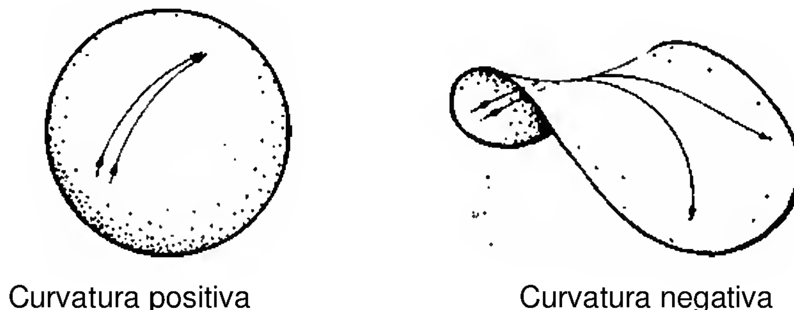


FIGURA V.28. *Geodésicas en una superficie curva. Con curvatura positiva las geodésicas convergen, mientras que con curvatura negativa divergen.*

La posibilidad de utilizar semejante descripción procede en última instancia de la hipótesis de Galileo (el principio de equivalencia) y nos permite eliminar la "fuerza" gravitatoria mediante la caída libre. De hecho, nada de lo que he dicho hasta ahora requiere que salgamos de la teoría newtoniana. Esta nueva imagen proporciona simplemente una *reformulación* de dicha teoría.[19] Sin embargo, la nueva física interviene cuando tratamos de combinar esta imagen con lo que aprendimos a partir de la descripción de Minkowski de la relatividad especial: la geometría del espacio-tiempo que ahora sabemos se aplica en *ausencia* de gravedad. La combinación resultante es la *relatividad general* de Einstein.

[19] La descripción matemática de esta reformulación de la teoría de Newton fue llevada a cabo por primera vez por el célebre matemático francés Élie Cartan (1923), lo que, por supuesto, ocurrió *después* de la teoría de la relatividad general de Einstein.

Recordemos lo que Minkowski nos ha enseñado. Tenemos (en ausencia de gravedad) un espacio-tiempo con un tipo peculiar de medida de "distancia" definida entre dos puntos: si tenemos una línea de universo en el espacio-tiempo, que describe la historia de alguna partícula, entonces la "distancia" minkowskiana medida a lo largo de la línea de universo describe el *tiempo* que dicha partícula experimenta realmente. (En realidad, en la sección previa considerábamos esta "distancia" sólo a lo largo de líneas de universo constituidas por tramos rectos, pero el enunciado se aplica también a líneas de universo curvas en las que la "distancia" se mide a lo largo de la curva.) La geometría de Minkowski se supone exacta si no hay campo gravitatorio, es decir, no hay curvatura del espacio-tiempo. Pero en presencia de gravedad consideramos la geometría de Minkowski como sólo aproximada —de la misma forma que una superficie plana da sólo una descripción aproximada de la geometría de una superficie curva—. Si imaginamos que tomamos un microscopio cada vez más potente para examinar una superficie curva —de modo que la geometría de la superficie aparezca cada vez más estirada— entonces la superficie aparece cada vez más plana. Decimos que una superficie curva es *localmente* semejante a un plano euclidiano.[20] De modo análogo, podemos decir que, en condiciones de gravedad, el espacio-tiempo es *localmente* semejante a la geometría de Minkowski (que es espacio-temporalmente *plana)* pero permitimos alguna curvatura en una escala mayor (véase fig. V.29).

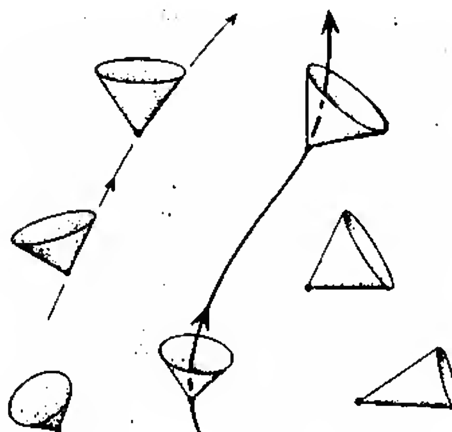


FIGURA V.29. *Representación del espacio-tiempo curvo.*

En particular, cualquier punto del espacio-tiempo es el vértice de un *cono de luz,* igual que en el espacio de Minkowski, pero estos conos de luz no están dispuestos de la manera uniforme en que lo estaban en el espacio de Minkowski. En el capítulo VII veremos algunos ejemplos de modelos de espacio-tiempo para los que esta no uniformidad es manifiesta *(cfr.* fig. VII.13, VII.14). Las partículas materiales tienen como líneas de universo curvas que están siempre dirigidas hacia el *interior* de los conos de luz, y los fotones tienen curvas dirigidas *a lo largo* de la superficie de los conos de luz. También hay un concepto de "distancia" minkowskiana a lo largo de tales curvas que mide el tiempo experimentado por las partículas, igual que en el espacio de Minkowski. Como en el caso de una superficie curva, esta medida de distancia define una *geometría* para la superficie, y puede diferir de la geometría del caso plano.

[20] Los espacios curvos localmente euclidianos en este sentido (también en mayores dimensiones) se llaman *variedades de Riemann* —en honor del gran Bernhard Riemann (1826-1866), quien los investigó por primera vez basándose en algunos importantes trabajos anteriores de Gauss para el caso bidimensional. Aquí necesitamos modificar la idea de Riemann, a modo de que la geometría pueda ser localmente *minkowskiana* en lugar de euclidiana. Tales espacios suelen denominarse *variedades lorentzianas* (y pertenecen a una clase llamada pseudo-riemanniana o, menos lógicamente, variedades semi-riemannianas)

Ahora puede darse a las geodésicas en el espacio-tiempo una interpretación similar a la de las geodésicas en superficies bidimensionales consideradas arriba, en donde debemos tener en cuenta las diferencias entre las situaciones minkowskiana y euclidiana. Así, en lugar de ser curvas de longitud mínima (localmente), nuestras líneas de universo geodésicas en el espacio-tiempo son curvas que *maximizan* (localmente) la "distancia" (es decir, el tiempo) a lo largo de la línea de universo. Las líneas de universo de las partículas en movimiento libre bajo gravedad *son* realmente geodésicas según esta regla. Así, en particular, los cuerpos celestes que se mueven en un campo gravitatorio están bien descritos por tales geodésicas. Además, los rayos de luz (líneas de universo de los fotones)

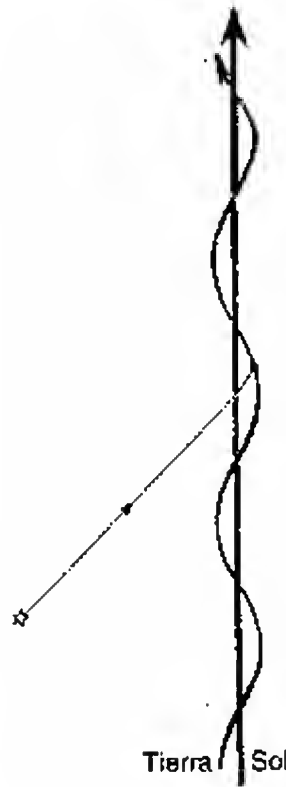


FIGURA V.30 *Líneas de universo de la Tierra y el Sol, y un rayo de luz procedente, de una estrella lejana que es desviado por el Sol.*

en el espacio vacío son también geodésicas, pero esta vez geodésicas de "longitud" *cero.*[21] Como ejemplo, he indicado esquemáticamente en la fig. V.30 las líneas de universo de la Tierra y el Sol, siendo el movimiento de la Tierra en torno al Sol una geodésica parecida a un sacacorchos en torno a la línea de universo del Sol. He indicado también un fotón que llega a la Tierra desde una estrella lejana. Su línea de universo aparece ligeramente "curvada" debido al hecho de que la luz, según la teoría de Einstein, es *desviada* por el campo gravitatorio del Sol.

Tenemos que ver aún cómo debe incorporarse la ley del inverso del cuadrado de Newton, y de qué modo debe modificarse de acuerdo con la relatividad de Einstein. Volvamos a nuestra esfera de partículas arrojadas en un campo gravitatorio. Recordemos que si la superficie esférica rodea

[21] Acaso preocupe al lector cómo este valor *cero* puede representar el valor *máximo* de la "longitud". De hecho lo hace, pero en un sentido vacuo: una geodésica de longitud cero se caracteriza por el hecho de que no existen líneas de universo de *ninguna otra* partícula que conecten cualquier par de sus puntos (localmente).

sólo un vacío, entonces, según la teoría de Newton, el volumen de la esfera inicial no cambia; pero si la superficie rodea materia de masa total *M*, entonces hay una reducción de volumen proporcional a *M*. En la teoría de Einstein las reglas son exactamente las mismas (para una esfera pequeña) salvo que no es exactamente *M* la que determina el cambio de volumen; hay una contribución (normalmente pequeñísima) de la *presión* en la materia rodeada. La expresión matemática completa para la curvatura del espacio-tiempo tetradimensional (que debe describir los efectos de marea para partículas que viajan en cualquier dirección posible en cualquier punto dado) viene dado por lo que se llama *tensor de curvatura de Riemann.* Este es un objeto algo complicado que requiere para su especificación veinte números reales en cada punto. Estos veinte números reales se denominan *componentes* del tensor. Las diferentes componentes se refieren a las diferentes curvaturas en diferentes direcciones del espacio-tiempo. El tensor de curvatura de Riemann se escribe normalmente R_{ijkl}, pero, puesto que no deseo explicar aquí qué significa cada uno de estos subíndices (ni, en realidad, qué *es* realmente un tensor), lo escribiré simplemente como:

RIEMANN

Este tensor puede descomponerse en dos partes, llamadas el tensor de *Weyl* y el tensor de *Ricci* (con diez *componentes cada uno).* Escribiré este desdoblamiento esquemáticamente como

RIEMANN = WEYL + RICCI.

(Las expresiones detalladas no nos serían especialmente útiles aquí.) El tensor de Weyl, que llamaremos WEYL, mide la *distorsión de la marea* de nuestra esfera de partículas en caída libre (es decir, un cambio inicial en forma, más que en tamaño), y el tensor de Ricci, que abreviaremos RICCI, mide su *cambio de volumen* inicial.[22] Recordemos que la teoría de la gravitación de Newton exige que la *masa* rodeada por nuestra superficie esférica en caída sea proporcional a esta reducción del volumen inicial. Esto nos dice, hablando en términos generales, que la densidad de *masa* de la materia —o equivalentemente la densidad de *energía* (puesto que $E = mc^2$)— debería *igualarse* al tensor de Ricci.

De hecho, esto es básicamente lo que afirman las ecuaciones de campo de la relatividad general, a saber, las *ecuaciones de campo de Einstein* [23]. Sin embargo, hay ciertas cuestiones técnicas sobre ello en las que será mejor que no entremos. Baste decir que existe un objeto llamado el tensor *energía-momento* que organiza toda la información pertinente acerca de la energía, presión y momento de la materia y los campos electromagnéticos. Me referiré a este tensor como ENERGÍA. Entonces las ecuaciones de Einstein se escriben, muy esquemáticamente:

RICCI = ENERGÍA

[22] En realidad esta división en efectos distorsionantes y cambio de volumen no es ni mucho menos tan clara como la he presentado. El mismo tensor de Ricci puede dar una cierta cantidad de distorsión de marea. (Con rayos de luz la división *es* completamente tajante (*cfr-* Penrose y Rindler, 1986, capítulo VII.) Una definición precisa de los tensores de Weyl y Ricci puede verse, por ejemplo, en Penrose y Rindler (1984, pp. 240, 210). (Hermann Weyl, de origen alemán, fue un matemático sobresaliente de este siglo; el italiano Gregorio Ricci fue un geómetra muy influyente que fundó la teoría de tensores en el siglo pasado.)

[23] La forma correcta de las auténticas ecuaciones fue encontrada también por David Hilbert en noviembre de 1915, pero las ideas físicas de la teoría se deben por entero a Einstein.

(Es la presencia de la "presión" en el tensor ENERGÍA, junto con algunos requisitos de consistencia para las ecuaciones en general, la que exige que la presión contribuya también al efecto de reducción de volumen arriba descrito.)

Esta ecuación no parece decir nada sobre el tensor de Weyl. No obstante, es una cantidad importante. El efecto de marea que se experimenta en el espacio vacío se debe enteramente a WEYL. De hecho, las ecuaciones de Einstein anteriores implican que existen ecuaciones *diferenciales* que ligan WEYL con ENERGÍA, algo similares a las ecuaciones de Maxwell que encontramos antes.[24] En realidad, un punto de vista fructífero consiste en considerar WEYL como un tipo de análogo gravitatorio de la cantidad campo electromagnético (también de hecho un tensor —el tensor de Maxwell) descrito por el par (E, B). Por consiguiente, en cierto sentido, WEYL mide el campo *gravitatorio*. La "fuente" de WEYL es el tensor ENERGÍA, lo que es análogo al hecho de que la fuente del campo electromagnético (E, B) es (*r, j*), el conjunto de cargas y corrientes de la teoría de Maxwell. Este punto de vista nos será útil en el capítulo VII.

Puede parecer curioso, cuando tenemos en cuenta estas sorprendentes diferencias en la formulación e ideas implícitas, que sea difícil descubrir diferencias observacionales entre la teoría de Einstein y la teoría que había desarrollado Newton dos siglos y medio antes. Pero siempre que las velocidades en consideración sean pequeñas comparadas con la velocidad de la luz *c,* y que los campos gravitatorios no sean demasiado intensos (de modo que las velocidades de escape sean mucho menores que *c; cfr.* capítulo VII), entonces la teoría de Einstein da resultados prácticamente idénticos a los de la de Newton. No obstante, la teoría de Einstein es más exacta en situaciones en las que *difieren* las predicciones de ambas teorías. Hay ahora varias verificaciones experimentales muy impresionantes, y la teoría más reciente de Einstein está completamente reivindicada. Los relojes marchan más lentos en un campo gravitatorio, como sostenía Einstein, habiendo sido medido este efecto de muchas maneras diferentes. La luz y las señales de radio son efectivamente desviadas por el Sol, y son ligeramente retardadas por este encuentro –efectos de relatividad general de nuevo perfectamente verificados.

Las sondas espaciales y los planetas en movimiento requieren pequeñas correcciones a las órbitas newtonianas, como exige la teoría de Einstein; éstas también han sido verificadas experimentalmente. (En particular, la anomalía en el movimiento del planeta Mercurio, conocida como la "precesión del perihelio", que había preocupado a los astrónomos desde 1859, fue explicada por Einstein en 1915.) Quizá lo más impresionante de todo sea un conjunto de observaciones de un sistema llamado el *pulsar binario,* consistente en un par de estrellas pequeñas de mucha masa (presumiblemente dos "estrellas de neutrones"), que concuerdan estrechamente con la teoría de Einstein y verifican indirectamente un efecto que está totalmente ausente en la teoría de Newton: la emisión de *ondas gravitatorias.* (Una onda gravitatoria es el análogo gravitatorio de una onda electromagnética, y viaja a la velocidad de la luz *c.)* No existe ninguna observación confirmada que contraiga la relatividad general de Einstein. A pesar de todo su extraño aspecto inicial, la teoría de Einstein sigue definitivamente vigente entre nosotros.

[24] Para quienes sepan de tales materias, estas ecuaciones diferenciales son las *identidades de Bianchi* en las que se han sustituido las ecuaciones de Einstein.

CAUSALIDAD RELATIVISTA Y DETERMINISMO

Recuérdese que en la teoría de la relatividad los cuerpos materiales no pueden viajar más rápido que la luz, en el sentido de que sus líneas de universo deben estar siempre dentro de los conos de luz *(cfr.* fig. V.29). (En relatividad general, particularmente, necesitamos expresar las cosas de este modo local. Los conos de luz no están dispuestos uniformemente, de modo que no tendría mucho sentido decir que aquí la velocidad de una partícula muy *lejana* supera la velocidad de la luz aquí.) Las líneas de universo de los fotones están *a lo largo* de las superficies de los conos de luz, pero ninguna partícula con líneas de universo puede *estar fuera* de los conos. De hecho debe satisfacerse un enunciado más general; a saber, que ninguna *señal* puede viajar fuera del cono de luz.

Para comprender por qué debe ser así, consideremos nuestra presentación del espacio de Minkowski (fig. V.31). Supongamos que se ha

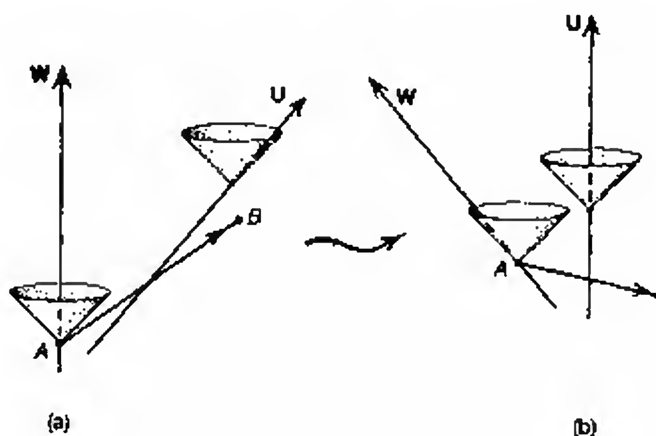


FIGURA V.31. *Una señal que para el observador W es más rápida que la luz parecerá, para el observador U, que viaja hacia atrás en el tiempo. La figura de la derecha* (b) *es simplemente la figura de la izquierda* (a) *redibujada desde el punto de vista de V. (Este nuevo dibujo puede imaginarse como un movimiento de Poincaré. Compárese con la fig. V.21, pero aquí, la transformación de* (a) *a* (b) *debe tomarse más en un sentido* activo *que pasivo.)*

construido algún dispositivo que pueda enviar una señal con velocidad un poco mayor que la de la luz. Utilizando este dispositivo el observador **W** envía una señal desde un suceso *A* de su línea de universo a un suceso distante *B* que está justo por debajo del cono de luz de *A*. En la fig. V.31a se ha dibujado esto desde el punto de vista de **W**, pero en la fig. V.31b esto se ha redibujado desde el punto de vista de un segundo observador **U** que se aleja rápidamente de **W** (digamos desde un punto entre *A* y *B),* y para el que el suceso *B* parece haber ocurrido *antes* que *A*. (Este "redibujado" es un movimiento de Poincaré, como el descrito arriba.) Desde el punto de vista de **W** los espacios simultáneos de **U** parecen estar "inclinados", que es por lo que el suceso *B* puede parecer a **U** anterior a *A*. Así, para **U**, la señal trasmitida por **W** parecería estar viajando hacia atrás en el tiempo.

Esto no es todavía una contradicción. Pero por simetría a partir del punto de vista de **U** (por el principio de relatividad especial), un *tercer* observador **V**, que se aleja de **U** en la dirección

opuesta de **W** y equipado con un dispositivo idéntico al de **W**, podría también enviar una señal un poco más rápida que la luz, desde *su* punto de vista (esto es, el de **V**) de retorno en la dirección de **U**. Esta señal podría parecer, para **U**, estar viajando también hacia atrás en el tiempo, ahora en la dirección espacial opuesta.

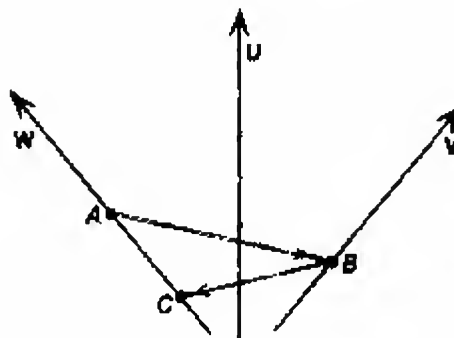


FIGURA V.32. *Sí V está equipado con un dispositivo que envía señales más rápidas que la luz, idénticas al de W, pero apuntando en la dirección opuesta, puede ser utilizado por W para enviar un mensaje a su propio pasado.*

En realidad **V** transmitirá esta segunda señal de retorno a **W** en el momento *(B)* que reciba la original enviada por **W**. Esta señal *alcanza* a **W** en un suceso *C* que es anterior, en la estimación de **U**, al suceso *A* de la emisión original (fig. V.32). Pero, lo que es aún peor, el suceso *C* es realmente anterior al suceso *A* de emisión en la *propia línea de universo* de **W**, de modo que **W** experimenta realmente que el suceso *C* ocurre antes de que él emita la señal en *A*. El mensaje que el observador **V** devuelve a **W** podría, por un acuerdo previo con **W**, repetir simplemente el mensaje que recibió en *B*. Así, **W** recibe, en un tiempo anterior en su línea de universo, precisamente el mismo mensaje que él va a enviar más tarde. Separando a los dos observadores a una distancia suficientemente grande, podemos disponer que la cantidad en la que la señal de retorno precede a la señal original sea un intervalo de tiempo tan grande como queramos. Quizá el mensaje original de W es que él se ha roto la pierna. Podría recibir el mensaje de retorno *antes* de que el accidente haya ocurrido y entonces (presumiblemente), por la acción de su libre voluntad, tomará las medidas para evitarlo.

Por lo tanto, las señales superlumínicas, junto con el principio de relatividad de Einstein, llevan a una patente contradicción con nuestras sensaciones normales de "libre albedrío". De hecho, la cuestión es aún más seria que eso. En efecto, podríamos imaginar que tal vez "el observador **W**" es simplemente un dispositivo mecánico programado para enviar el mensaje "sí" si recibe "no" y "no" si recibe "sí". De hecho, V puede ser también un dispositivo mecánico pero programado esta vez para enviar de retorno "no" si recibe "no" y "sí" si recibe "sí". Esto conduce a la misma contradicción esencial que teníamos antes,[25] ahora aparentemente independiente de la cuestión de si el observador **W** tiene o no "libre albedrío", y nos dice que un dispositivo de envío de señales más rápidas que la luz no es una "buena" posibilidad física. Esto tendrá más adelante para nosotros algunas implicaciones enigmáticas (capítulo VI).

Aceptemos entonces que *cualquier* tipo de señal —no simplemente las señales trasportadas por partículas físicas ordinarias— deben estar limitadas por los conos de luz. Si bien el argumento

[25] Hay algunas formas, no muy satisfactorias, de evitar este argumento *(cfr.* Wheeler y Feynman, 1945).

superior utiliza la relatividad *especial,* en la relatividad general las reglas de la relatividad especial son válidas localmente. Es la validez local de la relatividad especial la que nos dice que todas la señales están limitadas por los conos de luz, de modo que esto también se aplica en relatividad general. Veremos de qué forma afecta esto a la cuestión del *determinismo* en estas teorías. Recuérdese que en el esquema newtoniano (o en el hamiltoniano, etc.) "determinismo" significa que los *datos iniciales* en *un instante particular* fijan completamente el comportamiento para todos los demás instantes. Si adoptamos una visión espacio-temporal en la teoría newtoniana, entonces el "instante particular" en el que especificamos los datos sería alguna "rebanada" tridimensional a través del espacio-tiempo tetradimensional (es decir, la totalidad del espacio en dicho instante). En la teoría de la relatividad no existe un concepto global de "tiempo" que pueda seleccionarse para esto. El procedimiento usual consiste en adoptar una actitud más flexible. El "tiempo" de cualquiera es bueno. En relatividad especial podemos tomar algún espacio simultáneo de un observador en donde especificar los datos iniciales, en lugar de la "rebanada" anterior. Pero en relatividad general, el concepto de un "espacio simultáneo" no está muy bien definido. En su lugar podemos utilizar la noción más general de *superficie de tipo espacio.* [26] Una de estas superficies se representa en la fig. V.33; está caracterizada por el hecho de que está completamente fuera del cono de luz en cada uno de sus puntos, de modo que *localmente* se asemeja a un espacio simultáneo.

El determinismo, en relatividad especial, puede formularse como el hecho de que los datos iniciales en cualquier espacio simultáneo dado *S* fija el comportamiento en la totalidad del espacio-tiempo. (Esto será cierto, en particular, para la teoría de Maxwell, que es una teoría "relativista especial".) Sin embargo, aún podemos plantear un enunciado más fuerte. Si queremos saber lo que va a suceder en algún suceso P que está en alguna parte en el futuro de *S,* entonces sólo necesitamos los datos

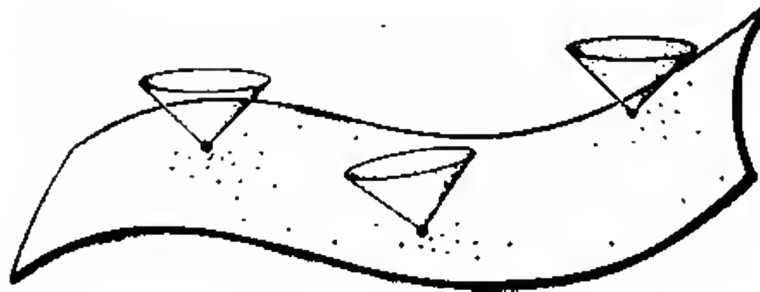


FIGURA V.33. *Superficie-espacio para la especificación de nuestros datos iniciales en relatividad general.*

[26] Técnicamente, el término "hipersuperficie" es más apropiado que "superficie" puesto que su espacio es tridimensional en lugar de bidimensional.



FIGURA V.34. *En relatividad especial, lo que sucede en P depende solamente de los datos en una región* finita *de un espacio simultáneo. Esto es debido a que los efectos no pueden viajar hacia P más rápido que la luz.*

iniciales en alguna región acotada (finita) de *S,* y no en la totalidad de S. Esto se debe a que la "información" no puede viajar más rápido que la luz, de modo que cualesquiera puntos de S que están demasiado alejados para que sus señales luminosas puedan alcanzar *P,* no pueden tener influencia en *P* (véase fig. V.34).* Esto es mucho más satisfactorio que la situación que surge en el caso newtoniano, en donde, en principio, podríamos tener necesidad de saber qué está pasando en toda la "rebanada" *infinita* para hacer cualquier predicción sobre lo que va a suceder en cualquier punto en un momento posterior. No existe límite para la velocidad a la que puede propagarse la información newtoniana y, de hecho, las fuerzas newtonianas son *instantáneas.*

El "determinismo" en relatividad *general* es un asunto mucho más complicado que en relatividad especial, y sólo haré algunos comentarios al respecto. En primer lugar, debemos utilizar una *superficie S de tipo espacio* para la especificación de datos iniciales (en lugar de una simple superficie simultánea). Entonces resulta que las ecuaciones de Einstein dan un comportamiento *localmente* determinista para el campo gravitatorio, suponiendo (como es usual) que los campos de materia que contribuyen al tensor ENERGÍA se comportan de forma determinista. Sin embargo, hay complicaciones considerables. La propia geometría del espacio-tiempo, incluyendo su estructura "causal" de conos de luz, es ahora parte de lo que hay que determinar. No conocemos esta estructura de conos de luz en el tiempo por venir, de modo que no podemos decir qué partes de S serán necesarias para determinar el comportamiento en algún suceso futuro *P.* En algunas situaciones extremas puede darse el caso de que incluso la S *entera* sea insuficiente y, consiguientemente, se pierde el determinismo global. (Aquí están implícitas cuestiones difíciles relacionadas con un importante problema no resuelto en la teoría de la relatividad general, llamado "censura cósmica", que tiene que ver con la formación de *agujeros negros* [Tipler y cols., 1980]; (*cfr.* capítulo VII). Podría parecer muy improbable que cualquiera de estos posibles "fallos de determinismo" que pudieran ocurrir con campos gravitatorios "extremos" tuviera una conexión directa con asuntos a la escala humana de las cosas, pero vemos a partir de esto que la cuestión del determinismo en relatividad general no es en absoluto tan clara como quisiéramos.

* Puede señalar que la ecuación de ondas es también, como las ecuaciones de Maxwell, una ecuación relativista. Por lo tanto, el "fenómeno de no computabilidad" de Pour-El y Richards que consideramos antes es un efecto que también se refiere solamente a datos iniciales en una región acotada de S.

LA COMPUTABILIDAD EN FÍSICA CLÁSICA: ¿DÓNDE ESTAMOS?

A lo largo de este capítulo he tratado de no perder de vista el tema de la *computabilidad,* como distinto al del determinismo, y he tratado de indicar que los temas de computabilidad pueden ser tan importantes, al menos, como los del determinismo cuando afectan a los cuestiones del "libre albedrío" y los fenómenos mentales. Pero el propio determinismo ha resultado no ser ni mucho menos tan claro, en la teoría clásica, como nos habían llevado a pensar. Hemos visto que la ecuación de Lorentz clásica para el movimiento de una partícula cargada da lugar a algunos problemas molestos. (Recuérdese las "soluciones desbocadas" de Dirac.)

Hemos señalado también que existen algunas dificultades para el determinismo en relatividad general. Cuando, en tales teorías, no existe determinismo tampoco hay ciertamente computabilidad. Pero en ninguno de los casos recién citados parece que la falta de determinismo tenga mucha importancia filosófica directa para nosotros. No hay aún "lugar" para el libre albedrío en dichos fenómenos: en el primer caso, debido a que la ecuación de Lorentz clásica para una partícula puntual (de la forma resuelta por Dirac) no se considera físicamente apropiada en el nivel en que aparecen estos problemas; y en el segundo, debido a que las escalas en que la relatividad general clásica podría llevar a tales problemas (agujeros negros, etc.) son completamente diferentes de las de nuestro cerebro.

Ahora bien, ¿dónde estamos con respecto a la *computabilidad* en la teoría clásica? Es razonable conjeturar que, con la relatividad general, la situación no es muy diferente de aquella en la relatividad especial, aparte de las diferencias en causalidad y determinismo que acabo de presentar. Allí donde el comportamiento futuro del sistema físico esté determinado a partir de datos iniciales, dicho comportamiento futuro parecerá estar (por razones similares a las que presenté en el caso de la teoría newtoniana) también *computablemente* determinado por estos datos[27] (aparte del tipo "inútil" de no computabilidad encontrado por Pour-El y Richards para la ecuación de ondas que consideramos arriba, y que no ocurre para datos que varían *suavemente).* En realidad, es difícil ver que en *cualquiera* de las teorías físicas presentadas hasta aquí pueda haber elementos "no computables" significativos. Con todo, debe esperarse que pueda ocurrir comportamiento caótico en muchas de estas teorías, en donde cambios muy pequeños en los datos iniciales pueden dar lugar a enormes diferencias en el comportamiento resultante. (Este parece ser el caso en relatividad general, *cfr.* Misner, 1969; Belinskii y cols., 1970.) Pero, como mencioné antes, es difícil ver cómo *este* tipo de no computabilidad —esto es, impredecibilidad— podría ser de "utilidad" en un dispositivo que trate de aprovechar los posibles elementos no computables en las leyes físicas. Si la "mente" puede estar haciendo uso de algún modo de los elementos no computables, entonces parece que deberían ser elementos que están fuera de la física clásica. Tendremos que revisar mas tarde esta cuestión, después de haber echado una mirada a la teoría cuántica.

MASA, MATERIA Y REALIDAD

Hagamos un breve inventario de esa imagen del mundo que nos ha presentado la física cuántica. En primer lugar, existe un espacio-tiempo que desempeña la primordial función de escenario en que se desenvuelve toda la diversa actividad de la física. En segundo lugar, existen *objetos*

[27] Sería de gran utilidad e interés contar con *teoremas* rigurosos sobre estos puntos. Por el momento faltan.

físicos entregados a esta actividad, aunque limitados por leyes matemáticamente precisas. Los objetos físicos son de dos tipos: *partículas* y *campos.* Poco se ha dicho sobre la naturaleza y las propiedades distintivas de las partículas, salvo que cada una tiene su propia línea de universo y posee una masa (en reposo) individual y tal vez carga eléctrica. Los campos, por otro lado, están muy específicamente dados: estando sujeto el campo electromagnético a las ecuaciones de Maxwell y el campo gravitatorio a las ecuaciones de Einstein.

Existe alguna ambigüedad sobre cómo hay que tratar a las partículas. Si tienen masas tan minúsculas que podemos soslayar su influencia sobre los campos, entonces las partículas se llaman *partículas de prueba,* y en este caso su movimiento en *respuesta* a los campos es inequívoco. La ley de fuerzas de Lorentz describe la respuesta de las partículas de prueba al campo electromagnético, y la ley geodésica describe su respuesta al campo gravitatorio (en una combinación apropiada, cuando ambos campos están presentes). Para esto, las partículas deben considerarse como partículas *puntuales,* es decir, que tienen líneas de universo unidimensionales. Sin embargo, cuando nos importan los efectos de las partículas sobre los campos (y, por lo tanto, sobre otras partículas) —es decir, cuando las partículas actúan como *fuentes* de los campos— entonces hay que considerarlas como objetos dispersos con cierta extensión. De otro modo los campos en el entorno inmediato de cada partícula se hacen infinitos. Estas fuentes extensas proporcionan la distribución de carga-corriente (x, j) que se necesita en las ecuaciones de Maxwell y el tensor ENERGÍA que se necesita para las ecuaciones de Einstein. Además de todo esto, el espacio-tiempo —donde residen todas las partículas y campos— tiene una estructura variable que por sí misma describe directamente la gravitación. El "escenario" interviene en la propia acción a la que sirve de marco.

Esto es lo que la física clásica nos ha enseñado sobre la realidad física. Es evidente que hemos aprendido mucho, aunque tampoco deberíamos confiar demasiado en que las imágenes que nos hemos formado no vayan a ser trastocadas por alguna visión posterior más penetrante. Veremos en el próximo capítulo que incluso los cambios revolucionarios que ha operado la teoría de la relatividad palidecen hasta resultar casi insignificantes en comparación con los de la teoría cuántica. No obstante, no hemos acabado aún con la teoría clásica y con lo que tiene que decirnos sobre la realidad material. Todavía nos reserva una sorpresa.

¿Qué es la "materia"? Es la sustancia real de la que están compuestos los objetos físicos, las "cosas" de este mundo. Es de lo que estamos hechos usted, yo y nuestras casas. ¿Cómo *cuantificamos* esta sustancia? Nuestros libros de texto de física elemental nos proporcionan la clara y nítida respuesta de Newton. Es la *masa* de un objeto, o de un sistema de objetos, la que mide la cantidad de materia que contiene ese objeto o sistema de objetos. Esto parece correcto, porque no hay ninguna otra cantidad física que pueda competir seriamente con la masa como la medida de sustancia total. Además se *conserva* la masa, y por consiguiente el contenido total de materia de un sistema cualquiera debe seguir siendo siempre el mismo.

Pero la famosa fórmula de la relatividad especial

$$E = mc^2$$

nos dice que masa *(m)* y energía *(E)* son intercambiables. Por ejemplo, cuando se desintegra un átomo de uranio, descomponiéndose en partes más pequeñas, la masa total de todas estas partes, si pudieran ser llevadas al reposo, sería *menor* que la masa original del átomo de uranio; pero si

se tiene en cuenta la *energía de movimiento* —energía *cinética,*[*]— de cada parte, y se convierte en valores de masa dividiéndola por c^2 (pues $E = mc^2$), entonces encontraremos que el total es *invariante.* La masa se conserva, pero al estar compuesta en parte de energía parece ahora menos evidente que sea la medida de la sustancia. Después de todo, la energía depende de la velocidad a que esté viajando la sustancia. La energía de movimiento de un tren expreso es considerable, pero si vamos sentados en el tren, según nuestro propio punto de visita el tren no tiene movimiento en absoluto. La energía de ese movimiento (aunque no así la *energía térmica* de los movimientos aleatorios de las partículas individuales del móvil) ha sido "reducida a cero" mediante esta apropiada elección del punto de vista. Para un ejemplo sorprendente, en donde el efecto de la relación masa-energía de Einstein es más extremo, consideremos la desintegración de cierto tipo de partícula subatómica llamada mesón π °. Ésta es ciertamente una partícula *material* que tiene una masa (positiva) bien definida. Después de alrededor de 10^{-16} segundos se desintegra (como el átomo de uranio anterior pero mucho más rápidamente) casi siempre en sólo *dos fotones* (fig. V.36)

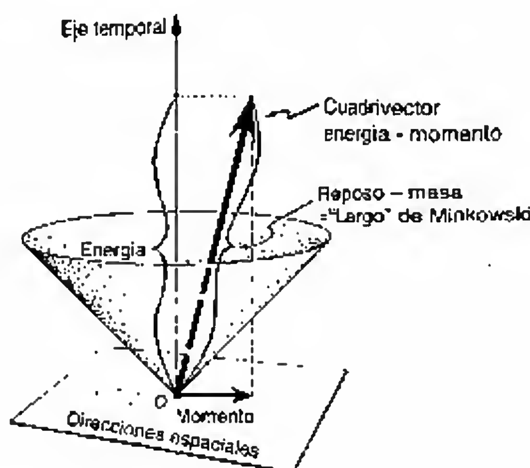


FIGURA V.35. *El cuadrivector energía-momento.*

[*] En la teoría newtoniana la energía cinética es 1/2 mv^2, donde *m* es la masa y *v* la velocidad; pero en la relatividad especial la expresión es algo mas complicada

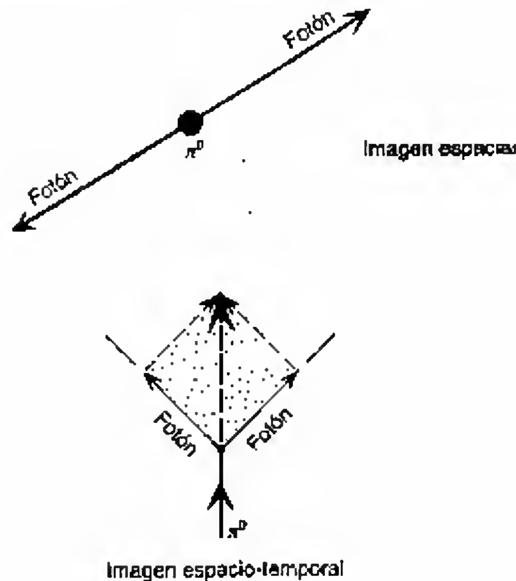


FIGURA V.36. *Un mesón π ° con masa se desintegra en dos fotones sin masa. La imagen espacio-temporal muestra cómo el cuadrivector energía-momento se conserva: el cuadrivector del mesón π ° es la suma de los cuadrivectores de los dos fotones sumados según la ley del paralelogramo (representado sombreado).*

Para un observador en reposo con el mesón π °, cada fotón lleva la mitad de la energía y, por lo tanto, la mitad de la masa del mesón π °. Pero esta "masa" del fotón es de un tipo vago: *pura energía.* En efecto, si tuviéramos que viajar rápidamente en la dirección de uno de los fotones, entonces podríamos reducir su masa-energía a un valor tan pequeño como quisiéramos, siendo *cero* realmente la masa intrínseca (o masa en *reposo como* veremos en breve) de un fotón. Todo esto forma una imagen consistente de la masa conservada, pero no es ya lo que teníamos antes. La masa puede aun, en cierto sentido, medir la "cantidad de materia", pero ha habido un cambio de punto de vista: puesto que la masa es equivalente a la energía, la masa de un sistema depende, como la energía, del movimiento del observador.

Merece la pena ser algo más explícitos acerca del punto de vista al que hemos desembocado. La cantidad conservada que asume el papel de la masa es un objeto global llamado el *cuadrivector energía-momento.* Este puede representarse como una flecha (vector) en el origen *O* del espacio de Minkowski que apunta *hacia el interior* del cono de luz futuro en *O* (o, en el caso extremo de un fotón, *sobre* este cono); véase fig. V.35. Esta flecha, que apunta en la misma dirección que la línea de universo del objeto, contiene toda la información sobre la energía, masa y momento. Así el "valor *t"* (o "altura") de la punta de esta flecha, medida en el sistema de referencia de algún observador, describe la *masa* (o *energía* dividida por c^2) del objeto, según dicho observador, mientras que las componentes especiales proporcionan el *momento* (dividido por *c).*

La "longitud" minkowskiana de esta flecha es una cantidad importante conocida como la *masa en reposo.* Describe la masa para un observador en reposo con el objeto. Podríamos tratar de adoptar el punto de vista de que *esto* sería una buena medida de la "cantidad de materia". Sin embargo no es aditiva: si un sistema se desdobla en dos, entonces la masa en reposo original no es la suma de las dos masas en reposo resultantes.

Recuérdese la desintegración del mesón π° considerada antes. El mesón π° tiene una masa en reposo positiva, mientras que las masas en reposo de cada uno de los dos fotones resultantes es nula. Sin embargo, la Propiedad de aditividad es satisfecha por la flecha global (cuadrivector), donde ahora debemos "sumar" en el sentido de la ley de la suma *vectorial* representada en la fig. V.6. Esta *flecha entera* será nuestra medida de la "cantidad de materia".

Consideremos ahora el campo electromagnético de Maxwell. Hemos Asistido en que lleva energía. Por $E = mc^2$, debe tener también masa. por lo tanto, el campo de Maxwell es también materia. Debemos aceptar esto ahora puesto que el campo de Maxwell está íntimamente relacionado con las fuerzas que ligan las partículas. Debe hacer una contribución sustancial[28] a la masa de cualquier cuerpo que procede del interior de su campo electromagnético.

¿Qué hay del campo gravitatorio de Einstein? En muchos aspectos se parece al campo de Maxwell. De modo análogo a como, en la teoría de Maxwell, los objetos cargados en movimiento pueden emitir ondas *electromagnéticas,* los cuerpos con masa en movimiento pueden emitir (según la teoría de Einstein) ondas *gravitatorias* —que, como las ondas electromagnéticas, viajan a la velocidad de la luz y transportan energía.

Pero esta energía no se mide de la manera estándar, que sería mediante el tensor ENERGÍA antes mencionado. En una onda gravitatoria (pura) el tensor es *cero* en cualquier parte. Podríamos aceptar, de todas formas, la idea de que la curvatura del espacio-tiempo (ahora dada completamente por el tensor WEYL) puede representar de algún modo la "sustancia" de las ondas gravitatorias. Pero la energía gravitatoria resulta ser *no local,* lo que quiere decir que no podemos determinar cuál es la medida de la energía examinando simplemente la curvatura del espacio-tiempo en regiones limitadas. La energía —y por consiguiente la masa— de un campo gravitatorio es un anguila escurridiza y se niega a quedarse quieta en ninguna localización clara. De todas formas, debe tomarse en serio. Está ciertamente *ahí,* y debe tenerse en cuenta para que el concepto de masa se conserve globalmente. Existe una buena medida de la masa (Bondi, 1960; Sanchs, 1962) que se aplica a las ondas gravitatorias, pero la no localidad es tal que hace que esta medida pueda a veces ser *no nula* en regiones *planas* del espacio-tiempo —entre dos estallidos de radiación (un poco como la calma en el ojo de un huracán)— en las que el espacio-tiempo *carece* totalmente de curvatura *(cfr.* Penrose y Rindler, 1986, p. 427); es decir, tanto WEYL como RICCI son nulos. En tales casos, parecemos obligados a deducir que si esta masa-energía tiene que estar localizada, debe estarlo en este *espacio vacío plano:* una región totalmente libre de materia o campos de cualquier tipo. En estas curiosas circunstancias, nuestra "cantidad de materia" o está *allí,* en las regiones más vacías del espacio vacío, o no estará en ninguna parte.

Esto parece una pura paradoja. Pero es clara consecuencia de lo que nos dicen nuestras mejores teorías clásicas —y son teorías realmente supremas— sobre la naturaleza de las sustancia "real" de nuestro mundo. La realidad material, según la teoría clásica, y ya no digamos en la teoría cuántica que estamos a punto de explotar, es algo mucho más vago de lo que hubiéramos imaginado. Su cuantificación e incluso su existencia o inexistencia— depende de puntos extraordinariamente sutiles y no puede verificarse localmente. Si esta localidad le parece enigmática, prepárese el lector para las sorpresas mucho mayores que vienen ahora.

[28] Incalculable, en la teoría actual y que hace inútil la respuesta (provisional): ¡infinito!

VI. MAGIA CUÁNTICA Y MISTERIO CUÁNTICO

¿NECESITAN LOS FILÓSOFOS LA TEORÍA CUÁNTICA?

EN FÍSICA CLÁSICA EXISTE, de acuerdo con el sentido común, un mundo objetivo "ahí fuera". El mundo evoluciona de un modo claro y determinista, gobernado por ecuaciones matemáticas formuladas exactamente. Esto es tan cierto para las teorías de Maxwell y Einstein como para el esquema original de Newton. Se considera que la realidad física existe independientemente de nosotros mismos; y el modo exacto de "ser" del mundo clásico no está afectado por el modo en que decidimos observarlo. Además, nuestros cuerpos y nuestros cerebros vienen a ser ellos mismos parte de este mundo. Se considera que también ellos evolucionan de acuerdo con las mismas ecuaciones clásicas exactas y deterministas. Todas nuestras acciones van a estar fijadas por estas ecuaciones; no importa cómo sintamos que nuestros deseos conscientes puedan influir en nuestro comportamiento.

Esta imagen parece estar en el fondo de los más serios argumentos filosóficos[1] concernientes a la naturaleza de la realidad, de nuestras percepciones conscientes y de nuestro aparente libre albedrío. Algunas personas no se quedarán tranquilas con esto y pensarán que también debería haber un papel para la *teoría cuántica,* ese fundamental pero perturbador esquema de cosas que surgió, en el primer cuarto de este siglo, a partir de las observaciones de discrepancias sutiles entre el comportamiento real del mundo y las descripciones de la física clásica.

Para muchos, el término "teoría cuántica" evoca simplemente el concepto vago de un "principio de incertidumbre" que, en el nivel de las partículas, átomos o moléculas, prohíbe la precisión en nuestras descripciones y da lugar a un comportamiento meramente probabilístico.

La verdad es que, como veremos, las descripciones cuánticas *son* muy precisas, aunque radicalmente diferentes de las descripciones clásicas y, pese a las opiniones en contra, las probabilidades *no* surgen en el ínfimo nivel cuántico de las partículas, átomos o moléculas —éstos evolucionan de modo *determinista*— sino que lo hacen por vía de cierta misteriosa acción, a mayor escala, ligada al surgimiento de un mundo clásico que podemos percibir conscientemente.

Debemos tratar de comprender esto y también de qué forma la teoría cuántica nos obliga a cambiar el concepto que tenemos de la realidad física. Se tiende a pensar que las discrepancias entre las teorías cuántica y clásica son muy insignificantes, pero de hecho involucran también a muchos fenómenos físicos a escala ordinaria: la existencia misma de los cuerpos sólidos, la resistencia y propiedades físicas de los materiales, la naturaleza de la química, los colores de las sustancias, los fenómenos de congelación y ebullición, la fiabilidad de la herencia... Éstas y muchas otras propiedades familiares requieren de la teoría cuántica para su explicación. El fenómeno de la conciencia es también algo que no puede entenderse en términos enteramente clásicos. Tal vez nuestras mentes son cualidades arraigadas en alguna extraña y misteriosa característica de las leyes físicas que gobiernan *realmente* el mundo en que vivimos, en lugar de

---

[1] He dado por supuesto que cualquier punto de vista filosófico "serio" debería contener al menos una buena dosis de realismo. ¡Siempre me sorprendo al enterarme de que *pensadores* aparentemente serios, con frecuencia físicos interesados en las implicaciones de la mecánica cuántica, adoptan la visión fuertemente subjetiva de que, en realidad, no existe en absoluto ningún mundo real "ahí fuera"! El hecho de que yo adopte una línea realista siempre que sea posible no quiere decir que no sea consciente de que tales visiones subjetivas se defienden seriamente muy a menudo, simplemente soy incapaz de entenderlas-Para un poderoso y divertido ataque a este subjetivismo, véase Gardner (1983), capítulo 1.

ser simples características de algún algoritmo ejecutado por los llamados "objetos" de una estructura física *clásica.* Quizá, en cierto sentido, ésta sea la "razón" de por qué debamos vivir, en tanto que seres sensibles, en un mundo cuántico en lugar de en uno enteramente clásico, a pesar de toda esa riqueza y de todo ese misterio que está presente en el universo clásico. ¿Sería *necesario* un mundo cuántico para que pudieran formarse, a partir de su sustancia, criaturas pensantes y perceptivas como nosotros mismos? Esta pregunta parece más apropiada para Dios, intentando construir un universo habitado, que para nosotros. Pero la pregunta también es pertinente para nosotros. Si la conciencia no puede formar parte de un mundo clásico, entonces nuestras mentes deben depender, de algún modo, de las desviaciones concretas respecto de la física clásica. Esta es una idea a la que volveré más adelante en este libro. Tenemos que entender la teoría cuántica —la más exacta y misteriosa de las teorías físicas— antes de ahondar en algunos temas mayores de la filosofía: ¿cómo se comporta nuestro mundo y qué es lo que constituye la individualidad de cada mente?

Algún día la ciencia podrá darnos acceso a una comprensión *más* profunda de la que nos proporciona por ahora la teoría cuántica. Mi opinión personal es que incluso la teoría cuántica es insuficiente e inadecuada para ofrecernos hoy una imagen completa acerca del mundo en el que realmente vivimos. Pero que esto no nos sirva de excusa. Es preciso que comprendamos la imagen del mundo según la teoría cuántica existente. Por desgracia, los teóricos tienden a tener enfoques muy diferentes (aunque desde puntos de observación similares) sobre la *realidad* de esta imagen.

Muchos físicos, encabezados por la figura señera de Niels Bohr, dirán que *no* hay imagen objetiva en absoluto. En el nivel cuántico nada hay realmente "ahí afuera". En cierto modo, la realidad emerge sólo en relación con los resultados de las "mediciones". Según esto, la teoría cuántica, proporciona simplemente un procedimiento de cálculo y no intenta describir el mundo como realmente "es". Esta actitud —hacia la teoría— me parece derrotista y por ello prefiero seguir la línea más positiva que atribuye una *realidad física objetiva a* la descripción cuántica: el *estado cuántico.*

Existe una ecuación muy precisa: la *ecuación de Schrödinger,* que proporciona una evolución temporal completamente determinista para este estado. Pero hay algo peculiar en torno a la relación entre el estado cuántico en su evolución temporal y el comportamiento real del mundo físico que se observa. De vez en cuando —siempre que consideremos que ha tenido lugar una "medición"— debemos descartar el estado cuántico que laboriosamente habíamos estado haciendo evolucionar y usarlo únicamente para calcular las diversas probabilidades de que el estado "salte" a uno u otro conjuntos de *nuevos* estados posibles. Además de lo extraño de ese "salto cuántico", está el problema de decidir cuál puede ser la configuración física que determina que realmente se ha hecho una "medición". Después de todo, el propio aparato de medición está construido presumiblemente a partir de componentes cuánticos y por eso debería evolucionar de acuerdo con la ecuación determinista de Schrödinger. Ahora bien ¿es necesaria la presencia de un ser consciente para que *realmente* tenga lugar una "medición"? Pienso que sólo una pequeña minoría de los físicos cuánticos sostendría esta opinión. Los observadores humanos están, ellos mismos, construidos a partir de minúsculos componentes cuánticos.

Más adelante reexaminaremos algunas de las extrañas consecuencias de este "salto" del estado cuántico —por ejemplo, una "medición" en un lugar origina un "salto" en una región lejana—. Antes de eso, encontraremos otro fenómeno extraño: a veces dos rutas opcionales —que un

objeto podría recorrer sin problemas, si cualquiera de ellas se atraviesa por separado— se cancelan completamente entre sí, cuando ambas están simultáneamente permitidas, de modo que en este último caso *ninguna* de ellas puede ser atravesada. Examinaremos también, con algún detalle, cómo se describen los estados cuánticos. Veremos cómo estas descripciones difieren mucho de sus correspondientes clásicas; por ejemplo, las partículas pueden parecer estar en dos lugares a la vez. También nos daremos una idea de cómo se complican las descripciones cuánticas al considerar varias partículas juntas (ya que éstas no tienen descripciones individuales para cada una de ellas, sino que deben considerarse como complicadas superposiciones de configuraciones alternativas de toda, ellas en conjunto). Veremos cómo distintas partículas del mismo tipo no pueden tener identidades separadas. Examinaremos en detalle la extraña (y fundamentalmente mecánico-cuántica) propiedad de *spin.* Consideraremos las importantes cuestiones que plantea el experimento mental paradójico del "gato de Schrödinger" y las diversas actitudes que han expresado los teóricos como intentos, en parte, de resolver este enigma fundamental.

Quizá parte del material de este capítulo no sea tan fácil de comprender como el de los capítulos precedentes (o los siguientes), y en ocasiones es un tanto técnico. He tratado de no hacer trampa en mis descripciones y tendremos que trabajar un poco más duro que en los otros casos. Esto es así para poder alcanzar una auténtica comprensión del mundo cuántico. Allí donde un argumento quede poco claro, le aconsejo que siga adelante y trate de hacerse una idea de la estructura global. Pero no desespere si una comprensión completa se le muestra esquiva. Es parte de la propia naturaleza de este tema.

PROBLEMAS CON LA TEORÍA CLÁSICA

¿Cómo sabemos que la física clásica no nos muestra la verdadera realidad de nuestro mundo? Las principales razones son experimentales. La teoría cuántica no era algo que desearan los teóricos; la mayoría de ellos fue conducida a su pesar, hacia esa extraña y, en muchos aspectos, filosóficamente insatisfactoria visión del mundo. Pero la teoría clásica, pese a su soberbia y su grandeza, entraña algunas dificultades, y la causa de éstas radica en la exigencia de que coexistan dos tipos de objetos físicos: las *partículas,* cada una de ellas descrita mediante un pequeño número finito (seis) de parámetros (tres posiciones y tres momentos); y los *campos,* que requieren un número *infinito* de parámetros.

Tal dicotomía no es físicamente consistente. Para que un sistema con partículas y campos esté en equilibrio (esto es, "completamente asentado"), toda la energía de las partículas debe cederse a los campos. Esta es una consecuencia del fenómeno llamado "equipartición de la energía": en el equilibrio la energía se reparte por igual entre todos los grados de libertad del sistema. Puesto que los campos tienen infinitos grados de libertad, a las pobres partículas no les queda nada en absoluto.

Del mismo modo, los átomos clásicos no serían estables porque todo el movimiento de las partículas se transferiría a los modos ondulatorios de los campos. Recordemos la imagen del átomo como un "sistema solar", introducida por el gran físico experimental anglo-neozelandés Ernest Rutherford en 1911. En lugar del Sol estaría el núcleo central, y en lugar de los planetas estarían los electrones en órbita —en una escala pequeñísima— mantenidos por el electromagnetismo en vez de la gravitación

Un problema fundamental y aparentemente insuperable es que cuan do un electrón orbital se moviera alrededor del núcleo debería emitir, de acuerdo con las ecuaciones de Maxwell, ondas electromagnéticas de una intensidad que crecería hasta infinito en una pequeñísima fracción de segundo, al tiempo que describiría una espiral que se cierra y se hunde en el núcleo. Sin embargo, nada de esto se observa sino algo que resulta inexplicable sobre la base de la teoría clásica. Los átomos pueden emitir ondas electromagnéticas (luz) pero sólo en destellos de discretas frecuencias específicas: las agudas *líneas espectrales* observadas (fig. VI. 1). Además, estas frecuencias satisfacen reglas "locas"[2] que no tienen ninguna base desde el punto de vista de la teoría clásica.

Otra manifestación que contradice los postulados sobre la coexistencia de campos y partículas es el fenómeno conocido como "radiación del cuerpo negro". Imaginemos un objeto a alguna temperatura definida, con la radiación electromagnética en equilibrio con las partículas. En 1900, Rayleigh y Jeans habían calculado que toda la energía sería absorbida, sin límite, por el campo. Hay un absurdo físico implícito en esto (la "catástrofe ultravioleta", cuando la energía sigue fluyendo sin cesar hacia el campo, con frecuencias cada vez mayores), y la propia naturaleza se comporta de forma más prudente. Para frecuencias *bajas* de las oscilaciones del campo, la energía es como habían predicho Rayleigh y Jeans, pero en el extremo de frecuencias *altas,* donde ellos habían predicho la catástrofe, las observaciones reales demuestran que la distribución de energía *no* crece sin límite sino que, por el contrario, cae a cero conforme crece la frecuencia. El mayor valor de la energía tiene lugar a una frecuencia (es decir, un color) muy concreta para una temperatura dada; véase fig. VI.2. (El rojo vivo de un atizador, o el color blanco-amarillento vivo del Sol son, en efecto, dos ejemplos familiares de esto.)

LOS COMIENZOS DE LA TEORÍA CUÁNTICA

¿Cómo iban a resolverse estos enigmas? El esquema original de Newton sobre las partículas *necesita,* ciertamente, ser complementado con el campo de Maxwell. ¿Podemos ir al otro extremo y suponer que *todo es* un campo, siendo las partículas pequeños nudos de tamaño finito de algún tipo de campo?

[2] En particular, J.J.Balmer señaló, en 1885, que las frecuencias de las líneas espectrales del hidrógeno tenían la forma $R(n^{-2} - m^{-2})$ en donde n y m son enteros positivos (siendo R una constante).

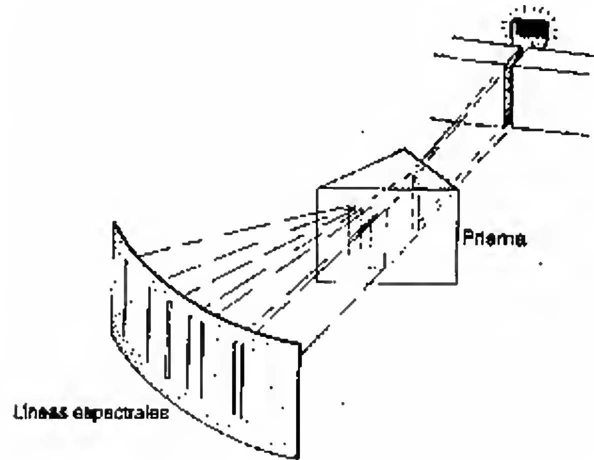


FIGURA VI. 1. Los *átomos en un material calentado emiten luz que a menudo resulta contener sólo frecuencias muy específicas. Las diferentes frecuencias pueden ser separadas mediante el uso de un prisma, y proporcionan las líneas espectrales características del átomo.*

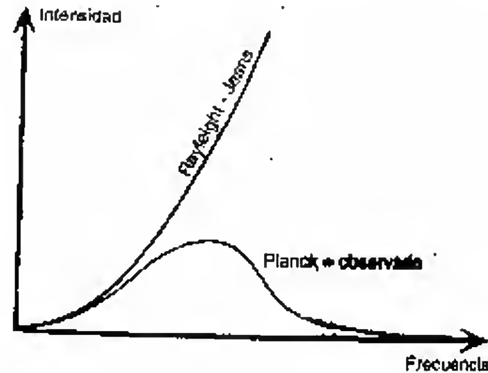


FIGURA VI. 2 *Las discrepancias entre la intensidad de la radiación calculada clásicamente Rayleigh-Jeans— y observada para un cuerpo caliente —"cuerpo negro"— llevaron a Planck a los inicios de la teoría cuántica.*

Esto también tiene sus dificultades, porque entonces las partículas podrían cambiar continuamente sus formas, agitándose y oscilando de innumerables y diferentes modos. Pero no es esto lo que vemos. En el mundo físico todas las partículas de la misma clase parecen ser *idénticas.* Por ejemplo, dos electrones cualesquiera son exactamente iguales entre sí, e incluso los átomos o las moléculas sólo pueden adoptar configuraciones de tipo discreto.[3] Si las partículas *fueran* campos entonces se necesitaría algún nuevo ingrediente para hacer posible que los *campos* adoptasen características discretas.

En 1900, el brillante aunque conservador y cauto físico alemán Max Planck propuso una idea revolucionaria para eliminar los modos de alta frecuencia del "cuerpo negro": que las

---

[3] Quizá no debiéramos descartar muy a la ligera esta imagen "enteramente de campos". Einstein, quien (como veremos) era profundamente consciente del carácter discreto que manifestaban las partículas cuánticas, dedicó los últimos treinta años de su vida a tratar de encontrar una teoría completamente global de este tipo clásico general. Pero los intentos de Einstein, como todos los demás, fracasaron. Parece necesitarse alguna otra cosa además de un campo clásico para explicar la naturaleza discreta de las partículas.

oscilaciones electromagnéticas sólo ocurren en "cuantos" cuya energía E mantiene una relación definida con la frecuencia ν , dada por

$$E = h\nu ,$$

siendo h una nueva constante fundamental de la naturaleza, hoy conocida como *constante de Planck.*

Con este ingrediente extravagante, Planck pudo obtener un sorprendente acuerdo teórico con la dependencia —experimentalmente observada— de la intensidad con la frecuencia, la actualmente llamada *ley de radiación de Planck.* (La constante de Planck es pequeñísima para los niveles cotidianos, alrededor de 6.6×10^{-34} julios por segundo.)

Con este golpe maestro, Planck revelaba los albores de la teoría cuántica venidera, pero mereció escasa atención sólo hasta que Einstein hizo otra propuesta insólita: el campo electromagnético únicamente puede *existir* en estas unidades discretas. Recordemos que Maxwell y Hertz habían demostrado que la *luz* consiste en oscilaciones del campo electromagnético. Por lo tanto, según Einstein —y como Newton había insistido dos siglos antes— la propia luz debe estar compuesta, después de todo, de *partículas.* (A comienzos del siglo XIX, el brillante teórico y experimentador inglés Thomas Young había establecido aparentemente que la luz consistía en ondas.)

¿Cómo es posible que la luz pueda consistir en partículas y en oscilaciones del campo al mismo tiempo? Estas dos concepciones parece opuestas, pero algunos hechos experimentales indicaban que la luz es compuesta por partículas y otros que lo está por ondas. En 1923, el príncipe francés y perspicaz físico, Louis de Broglie, llevó esta confusión partícula-onda un paso más allá, cuando en su disertación doctoral (para la que buscó la aprobación de Einstein) propuso que las propias partículas de *materia se* comportarían a veces como ondas. La frecuencia de la onda de De Broglie: ν , para una partícula de masa m, satisface de nuevo la relación de Planck. Combinada con la relación de Einstein $E = mc^2$, esto nos dice que ν está relacionada con m mediante

$$h\nu = E = mc^2.$$

Así, según la propuesta de De Broglie, la dicotomía entre partículas y campos, que había sido una característica de la teoría clásica, *no* se respeta en la naturaleza. En realidad, cualquier cosa que oscile con alguna frecuencia v puede ocurrir *sólo* en unidades discretas de masa $h\nu /c^2$. La naturaleza consigue, de alguna manera, construir un mundo consistente en el que *partículas y oscilaciones de campo son la misma cosa.* O, más bien, su mundo está constituido por algún ingrediente más sutil, siendo las palabras "partícula" y "onda" imágenes sugerentes pero sólo parcialmente apropiadas.

Otra brillante utilización de la relación de Planck fue dada (en 1913) por Niels Bohr, físico danés y figura sobresaliente del pensamiento científico del siglo XX. Las reglas de Bohr exigían que el *momento angular* de los electrones en órbita en torno al núcleo pueda darse sólo en múltiplos enteros de $h/2\pi$, para lo que Dirac introdujo más tarde el símbolo conveniente $\hbar$

$$\hbar = \frac{h}{2\pi}$$

Por lo tanto, los únicos valores permitidos del momento angular (respecto a cualquier eje) son

$$0, h, 2h, 3h, 4h....$$

Con este nuevo ingrediente el modelo de "sistema solar" para el átomo daba ahora con una precisión considerable, muchos de los niveles de energía estables discretos, así como las reglas "locas" para las frecuencias espectrales que la naturaleza obedece *realmente.*

Aunque sorprendentemente acertada, la brillante propuesta de Bohr, proporciona un esquema de piezas sueltas algo provisional, conocido como la "antigua teoría cuántica".

La teoría cuántica como la conocemos hoy surgió de dos esquemas básicos independientes, los cuales fueron iniciados por un par de físicos extraordinarios: un alemán, Werner Heisenberg, y un austríaco, Erwin Schrödinger. Al principio estos dos esquemas (la "mecánica matricial" en 1925 y la "mecánica ondulatoria" en 1926) parecían bastante diferentes, pero pronto se demostró que eran equivalentes y fueron englobados en un marco más comprensivo y general, principalmente por obra del gran físico teórico británico Paul Adrien Maurice Dirac. En las secciones siguientes echaremos un vistazo a esta teoría y sus extraordinarias explicaciones.

EL EXPERIMENTO DE LA DOBLE RENDIJA

Consideremos el experimento mecánico-cuántico "arquetípico" y según el cual un haz de electrones, o de luz, o de algún otro tipo de "onda-partícula", se lanza a través de un par de estrechas rendijas hacia una pantalla (fig. VI.3). Para ser precisos, consideraremos la *luz* y nos referiremos a los cuantos de luz como "fotones", de acuerdo con la terminología usual. La manifestación más evidente de la luz como *partículas* (es decir, como fotones) tiene lugar en la pantalla. La luz llega allí en unidades de energía discretas y localizadas, estando esa energía invariablemente relacionada con la frecuencia de la luz, de acuerdo con la fórmula de Planck: $E=h\nu$. Nunca se recibe la energía de sólo "medio" fotón (o cualquier otra fracción). La recepción de la luz es un fenómeno de *todo o nada* en unidades de fotones. Solamente se han visto números enteros de fotones.

Sin embargo, parece surgir un comportamiento de tipo ondulatorio cuando los fotones atraviesan las rendijas. Supongamos, primero, que sólo está abierta una rendija (mientras la otra está bloqueada). Después de atravesarla, la luz se dispersará por el fenómeno llamado *difracción,* una característica de la propagación de ondas. Pero todavía podemos mantener una imagen de partícula e imaginar que la proximidad de los bordes de la rendija ejerce alguna influencia para que los fotones sean desviados a uno u otro lado en una cantidad aleatoria.

Cuando la luz que atraviesa la rendija tiene una intensidad razonable, es decir un gran número de fotones, la iluminación en la pantalla aparece muy uniforme,

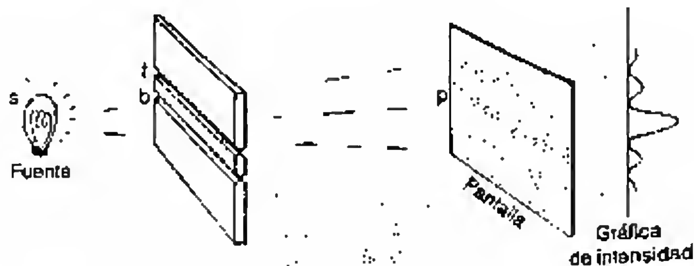


FIGURA VI.3. *El experimento de la doble rendija con luz monocromática.*

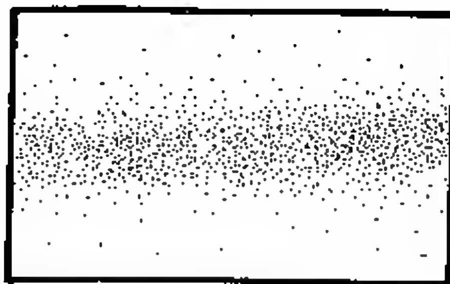


FIGURA VI.4. *Gráfica de intensidad en la pantalla cuando sólo está abierta una rendija: una distribución de minúsculos punto discretos.*

pero si se reduce la intensidad, entonces podemos concluir que la distribución de la iluminación está formada por puntos individuales —de acuerdo con la imagen de partículas— en donde los fotones individuales inciden sobre la pantalla. La apariencia tenue de la iluminación es un efecto estadístico debido al número muy grande de fotones involucrados (véase fig. VI.4). (Para que sirva de comparación, una bombilla de 60 watts emite alrededor de 100 000 000 000 000 000 000 fotones por segundo.) Los fotones parecen ser desviados efectivamente de una manera aleatoria cuando atraviesan la rendija, con probabilidades distintas para diferentes ángulos de desviación, produciendo la distribución de iluminación observada.

Pero, el problema clave para la imagen de partículas viene cuando abrimos la otra rendija. Supongamos que la luz procede de una lámpara amarilla de sodio, de modo que esencialmente consta de un color puro sin mezcla. El término técnico es *monocromático* —es decir, de una longitud de onda o frecuencia definida—, lo que significa, en la imagen de la partícula, que todos los fotones tienen la misma energía. En este caso, la longitud de onda es de unos 5×10^{-7} m.

Tomemos las rendijas de aproximadamente 0.001 mm de anchura y separadas unos 0.15 mm, con la pantalla a un metro de distancia. Para una intensidad de luz razonablemente alta seguimos teniendo una figura de iluminación de apariencia regular, pero ahora hay una *ondulatoriedad* en ella, llamada "figura de interferencia", con bandas de unos tres milímetros de anchura cerca del centro de la pantalla (fig. VI.5).

Hubiéramos esperado que la apertura de la segunda rendija simplemente duplicaría la intensidad de iluminación en la pantalla. De hecho, esto es cierto si consideramos la iluminación *total* en conjunto. Pero ahora la figura detallada de intensidad se ve completamente distinta de la que era con una sola rendija. En algunos puntos de la pantalla — en donde la figura alcanza su mayor brillo— la intensidad de iluminación es el *cuádruple* de la que era antes, y no sólo el doble. En otros—donde la figura es más oscura— la intensidad se hace nula.

Los puntos de intensidad nula plantean quizá el mayor enigma para la imagen en términos de partícula. Aquellos eran puntos a los que podía llegar un fotón con bastante facilidad cuando sólo una de las rendijas estaba abierta. Ahora que hemos abierto la otra, resulta repentinamente que el fotón está *imposibilitado* para hacer lo que podía hacer antes ¿Cómo es posible que, al permitir una ruta *alternativa* al fotón, el efecto real sea que le hayamos *impedido* atravesar cualquiera de las rutas?

Si tomamos la longitud de onda del fotón como una medida, la segunda rendija estará separada unos 300 "tamaños de fotón" de la primera, y cada rendija tendrá de ancho un par de longitudes

de onda (véase fig. VI.6). Entonces, ¿cómo puede "saber" el fotón cuando atraviesa una de las rendijas, si la otra rendija está o no abierta? De hecho, en principio no hay límite para la distancia a la que pueden estar las dos rendijas antes de que ocurra este fenómeno de "cancelación" o de "refuerzo".

Cuando la luz atraviesa la(s) rendija(s), parece estar comportándose como una *onda* y *no* como una partícula. Semejante cancelación —*interferencia destructiva*— es una propiedad familiar de las ondas ordinarias. Si una onda puede viajar por una u otra de las dos rutas y si *ambas* le son accesibles, entonces es posible que ellas se cancelen mutuamente. En la fig. VI.7 he ilustrado cómo sucede esto. Cuando una porción de la onda que ha atravesado una de las rendijas encuentra una porción que ha atravesado la otra, ambas se refuerzan mutuamente si están "en fase" (es decir, cuando coinciden las crestas de las dos porciones y también coinciden sus valles), pero se cancelan una a la otra si están exactamente "fuera de fase" (lo que significa que una tiene una cresta donde la otra tiene un valle).

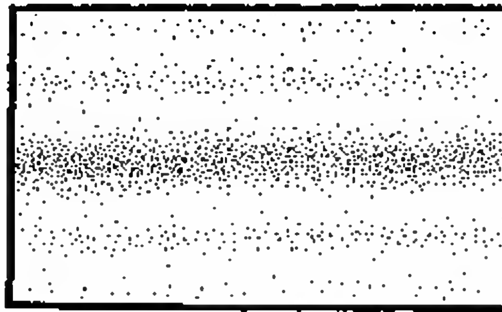


FIGURA VI.5. *Figura de intensidad, cuando ambas rendijas están abiertas, una distribución ondulada de puntos discretos.*

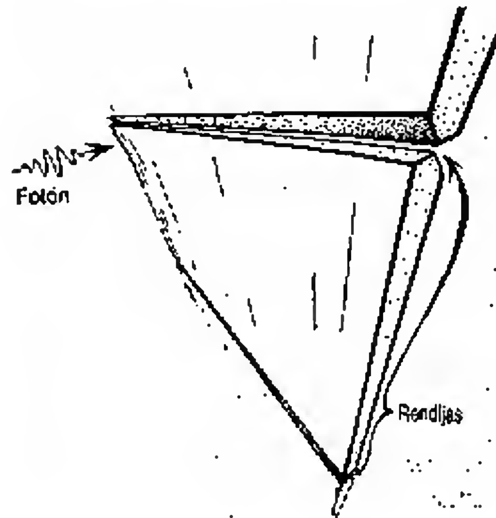


FIGURA VI.6. *Las rendijas "desde el punto de vista del fotón". ¿Cómo puede suponer una diferencia para él que la segunda rendija, a una distancia de unos 300 "tamaños de fotón", esté abierta o cerrada?*

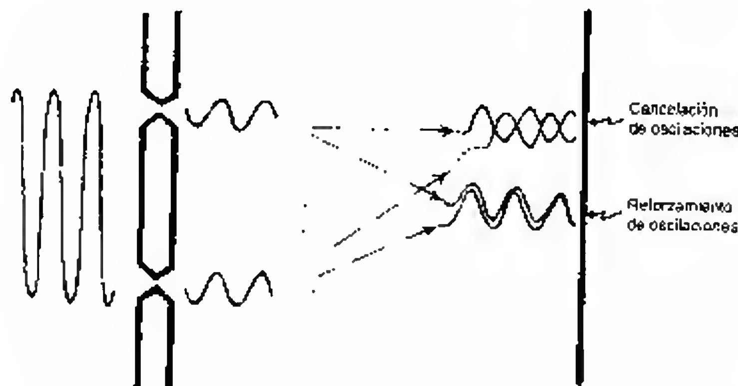


FIGURA VI.7. *Con una imagen puramente de ondas podemos entender la figura de bandas brillantes y oscuras en la pantalla —aunque no su carácter discreto— en términos de interferencia de ondas.*

Con el experimento de la doble rendija, los lugares brillantes de la pantalla aparecen siempre y cuando las distancias a las dos rendijas difieran en un número *entero* de longitudes de onda, de modo que coincidan las crestas y valles. Y los lugares oscuros se dan cuando las diferencias entre las dos distancias están exactamente a medio camino entre estos valores, de modo que cresta coincide con valle y valle con cresta.

No hay nada enigmático en una onda clásica macroscópica ordinaria que al mismo tiempo viaja a través de las dos rendijas. Después de todo, una onda es sólo una "perturbación", bien de algún medio continuo (campo) o bien de alguna sustancia compuesta de miríadas de minúsculas partículas puntuales.

Una perturbación puede atravesar en parte una rendija y en parte la otra. Pero ahora las cosas son muy diferentes: cada fotón individual se comporta por sí mismo enteramente como una onda. En cierto sentido, cada partícula atraviesa *ambas rendijas a la vez* e interfiere *consigo misma.* Reduciendo suficientemente la intensidad global de la luz podemos asegurar que no más de un fotón está cada vez en la vecindad de las rendijas. El fenómeno de la interferencia destructiva, en el que se realiza la posibilidad de que dos rutas opcionales para el fotón logran de algún modo cancelarse mutuamente, es algo que se aplica a fotones *individuales.* Si sólo una de las dos rutas está abierta al fotón, entonces el fotón puede recorrer la ruta. Si sólo la otra ruta está abierta, entonces el fotón puede recorrer esta otra ruta en lugar de la anterior. Pero si *ambas* están abiertas las dos posibilidades se cancelan y el fotón parece incapaz de recorrer cualquiera.

El lector debería detenerse a considerar la importancia de este hecho extraordinario. No se trata en realidad de que la luz se comporte a veces como partículas y a veces como ondas. Se trata de que *cada partícula individual* se comporta por sí misma de una manera ondulatoria; y de que *diferentes posibilidades abiertas a una partícula pueden cancelarse mutuamente.*

¿Realmente se desdobla el fotón, de tal modo que parte de él atraviesa una rendija y otra parte atraviesa la otra? La mayoría de los físicos objetaría esta forma de expresar las cosas. Insistirían en que, si bien las dos rutas abiertas a la partícula deben contribuir al efecto final, éstas son precisamente rutas *opcionales* y no debe —por lo tanto— suponerse que la partícula se desdobla en dos al atravesar las rendijas.

En apoyo de esta opinión de que la partícula no atraviesa en parte una rendija y en parte la otra, puede considerarse la situación modificada en la que un *detector de partículas* se coloca en una u otra de las rendijas.

Puesto que cuando se observa un fotón —o cualquier otra partícula— aparece siempre como un todo individual, y no como una fracción de un todo, nuestro detector debería o bien detectar un fotón entero o bien no detectar nada en absoluto. Sin embargo, cuando el detector está presente en una de las rendijas de modo que el observador puede *decir* a través de cuál rendija ha pasado el fotón, la figura de interferencia ondulatoria en la pantalla desaparece.

Aparentemente, para que tenga lugar la interferencia debe haber una "falta de conocimiento" acerca de qué rendija atravesó "realmente" la partícula. Para obtener interferencia, ambas opciones deben contribuir, a veces "sumándose" —reforzándose mutuamente en una cantidad que es el doble de lo que cabría esperar— y a veces "restándose", de modo que las opciones puedan "cancelarse" mutuamente. En realidad, según las reglas de la mecánica cuántica, lo que está sucediendo es todavía más misterioso que eso. Las opciones pueden efectivamente sumarse (los puntos más brillantes de la pantalla) y restarse (los puntos oscuros), pero también deben poder combinarse en otras formas extrañas, tales como:

"alternativa A" más i x "alternativa B",

donde *"i"* es la "raíz cuadrada de menos uno" ($\sqrt{-1}$) que encontramos en el capítulo III (en los puntos de la pantalla de intensidad intermedia).

De hecho, *cualquier número complejo* puede jugar ese papel en la "combinación de opciones". El lector recordará mi advertencia en el capítulo III acerca de que los números complejos son "absolutamente fundamentales" para la estructura de la mecánica cuántica. Tales números no son simples sutilezas matemáticas, y se han abierto paso hasta la atención de los físicos a través de persuasivos e inesperados hechos experimentales.

Para comprender la mecánica cuántica debemos entender las ponderaciones o pesos estadísticos complejos. A continuación consideraremos lo que esto implica.

AMPLITUDES DE PROBABILIDAD

No hay nada especial en la utilización de fotones para las descripciones anteriores. Servirían igualmente los electrones o cualquier otro tipo de partículas, o incluso átomos enteros. Las reglas de la mecánica cuántica parecen incluso insistir en que las bolas de cricket y los elefantes deberían comportarse de esta manera singular, según la cual diferentes Posibilidades opcionales pueden "sumarse" de algún modo en combinaciones de números complejos. Sin embargo, nunca *vemos* bolas de cricket o elefantes superpuestos de esta extraña manera. ¿Por qué no los vemos? Este es un tema difícil e incluso controvertido, y no quiero abordarlo todavía. Por el momento, como hipótesis de trabajo, supongamos simplemente que existen dos posibles niveles de descripción física: el *nivel cuántico* y el *nivel clásico.* Sólo utilizaremos estas extrañas combinaciones de números complejos en el nivel cuántico. Las bolas de cricket y los elefantes son objetos del nivel clásico.

El nivel cuántico es el nivel de las moléculas, átomos, partículas subatómicas, etc. Normalmente se suele pensar que es un nivel de fenómenos de muy "pequeña escala", pero esta "pequeñez" se

refiere en realidad al tamaño físico. Veremos que los efectos cuánticos pueden ocurrir a distancia de muchos metros, o incluso años luz. Estaría mucho más cerca de la realidad pensar que algo está "en el nivel cuántico" si implica sólo diferencias muy pequeñas de energía. (Trataré de precisarlo más adelante, especialmente en el capítulo VIII.). El nivel clásico es el nivel "macroscópico", del que somos más directamente conscientes. Es un nivel en el que nuestra imagen ordinaria de las "cosas que suceden" se mantiene válida, y en el que podemos utilizar nuestras ideas ordinarias de probabilidad.

Veremos que los números complejos que debemos utilizar en el nivel cuántico están estrechamente relacionados con las probabilidades clásicas. No son exactamente lo mismo, pero para entender estos números complejos será útil recordar, primero, cómo se comporta la probabilidad clásica. Consideremos una situación clásica que es *incierta,* de modo que no sabemos cuál de las dos opciones A o B tiene lugar. Semejante situación podría describirse mediante una combinación "ponderada" de estas opciones:

p x "alternativa A" más q x "alternativa B" ,

donde p es la *probabilidad* de que suceda A y q es la probabilidad de que suceda B. (Recuérdese que la probabilidad es un número real entre 0 y 1. Probabilidad 1 significa "certeza de que el fenómeno tendrá lugar" y probabilidad 0 significa "certeza de que el fenómeno no tendrá lugar". Probabilidad 1/2 significa "igualmente probable que tenga lugar y que no lo tenga".) Si A y B son las *únicas* opciones entonces la suma de las dos probabilidades debe ser 1:

$$p + q = 1.$$

Sin embargo, si hay más opciones esta suma debe ser menor que 1. Entonces, la *razón* $p:q$ da la *razón* de la probabilidad de que suceda A a la de que suceda B. Las probabilidades reales de que suceda A y de que suceda B —de entre estas dos únicas opciones— serán $p/(p+q)$ y $q/(p+q)$ respectivamente. Podríamos utilizar también la misma interpretación en el caso de que $p + q$ sea mayor que 1. (Esto podría ser útil, por ejemplo, en el caso en que tengamos un experimento que se realiza muchas veces, siendo p el número de ocurrencias de A y q el número de ocurrencias de B.) Diremos que p y q están *normalizadas* si $p + q = 1$, de modo que dan las probabilidades reales y no sólo la razón de esas probabilidades.

En física cuántica tenemos algo que se *parece* mucho a esto, excepto que p y q son ahora números *complejos,* que preferiré denotar, en su lugar, por w y z, respectivamente.

w x "alternativa A" más z x "alternativa B" .

¿Cómo tenemos que interpretar w y z? Ciertamente, no son probabilidades (o razones de probabilidades) ordinarias, puesto que cada una de ellas puede ser, independientemente, negativa o compleja, pero en muchos aspectos se comportan como probabilidades. Se las denomina —cuando están adecuadamente normalizadas (véase más adelante)— *amplitudes de probabilidad,* o simplemente *amplitudes.* Además, se utiliza frecuentemente el tipo de terminología que sugieren las probabilidades, tal como: "existe una amplitud w de que suceda A y una amplitud z de que suceda B". *No* son realmente probabilidades, pero por el momento pretenderemos que lo son o, mejor dicho, que son las análogas de las probabilidades en el nivel cuántico.

¿Cómo funcionan las probabilidades *ordinarias*? Será útil que pensemos en un objeto macroscópico, digamos una bola que es golpeada y lanzada (a través de uno de dos agujeros)

hacia una pantalla colocada atrás —como es el experimento de la doble rendija descrito antes (*cfr.* fig. VI.3)—, pero en donde ahora una bola macroscópica clásica reemplaza al fotón de la discusión previa. Habrá alguna probabilidad $P(f,s)$ de que la bola llegue al agujero superior s después de ser golpeada en f, y alguna probabilidad $P(f,i)$ de que llegue al agujero inferior i. Y, si seleccionamos un punto particular p en la pantalla, habrá alguna probabilidad $P(s,p)$ de que cuando la bola *pase* a través de s llegue al punto particular p de la pantalla, y alguna probabilidad $P(i,p)$ de que cuando pase a través de i llegue a p. Si solamente está abierto el agujero superior s, entonces para obtener la probabilidad de que la bola llegue a p vía s después de ser golpeada multiplicamos la probabilidad de que vaya *de* f a s por la probabilidad de que vaya de s a p:

$$P(f,s) \text{x} P(s,p)$$

Análogamente, si sólo está abierto el agujero inferior, entonces la probabilidad de que la bola vaya de f a p es

$$P(f,i) \text{x} P(i,p).$$

Si *ambos* agujeros están abiertos, entonces la probabilidad de que vaya de f a p vía s sigue siendo la primera expresión $P(f,s)$ x $P(s,p)$, igual que si sólo s estuviera abierto, y la probabilidad de que vaya de f a p vía i sigue siendo $P(f,i)$ x $P(i,p)$, de modo que la probabilidad *total* $P(f,p)$ de que llegue a p partiendo de f es la suma de estas dos:

$$P(f,p) = P(f,s) \text{ x } P(s,p) + P(f,i) \text{ x } P(i,p).$$

En el nivel *cuántico* las reglas son exactamente las mismas, excepto que ahora son estas extrañas *amplitudes* complejas las que deben desempeñar el papel de las probabilidades que teníamos antes. Así, en el ejemplo de la doble rendija antes considerado, tenemos una amplitud $A(f,s)$ de que un fotón llegue a la rendija superior s a partir de la fuente f y una amplitud $A(s,p)$ de que llegue al punto p de la pantalla a partir de la rendija s, y multiplicamos estas dos para obtener la amplitud

$$A(f,s) \text{ x } A(s,p),$$

de que llegue a la pantalla en p vía s. Como sucede con las probabilidades, esta es la amplitud correcta si suponemos que la rendija superior está abierta (esté abierta o no la rendija inferior i.) Análogamente, suponiendo que i está abierta, existe una amplitud

$$A(f,i) \text{ x } A(i,p),$$

de que el fotón llegue a p procedente de f vía i (esté o no abierta s). Si ambas rendijas están abiertas, entonces tenemos una amplitud total

$$A(f,p) = A(f,s) \text{ x } A(s,p) + A(f,i) \text{ x } A(i,p),$$

de que el fotón llegue a p procedente de f.

Todo esto está muy bien, pero no nos servirá de mucho hasta que sepamos cómo interpretar estas amplitudes cuando un efecto cuántico se amplifica y alcanza el nivel clásico. Podemos, por ejemplo, tener un detector de fotones, o *fotocélula,* colocada en p, que proporcione una manera de amplificar un suceso en el nivel cuántico —la llegada de un fotón a p— hasta un acontecer clásico discernible, digamos un "click" audible. (También valdría que la pantalla actuara como placa fotográfica, de modo que el fotón deje una mancha visible, pero por claridad nos

atendremos al uso de la fotocélula.) Debe haber entonces una *probabilidad* real de que ocurra el "click", y no sólo una de estas misteriosas "amplitudes". ¿Cómo tenemos que pasar de las amplitudes a las probabilidades cuando ascendemos desde el nivel cuántico al clásico? Resulta que para esto hay una regla muy bella aunque misteriosa.

La regla dice que debemos tomar el *módulo al cuadrado* de la amplitud compleja cuántica para obtener la probabilidad clásica. ¿Qué es un "módulo al cuadrado"? Recuérdese nuestra descripción de los números complejos en el plano de Argand (capítulo III). El *módulo* $|z|$ de un número complejo z es simplemente la distancia desde el origen (esto es, desde el punto 0) hasta el punto definido por z. El módulo al cuadrado $|z|^2$ es simplemente este módulo elevado al cuadrado. Así, si

$$z=x+iy,$$

donde x e y son números reales, entonces (por el teorema de Pitágoras, ya que la línea de 0 a z es la hipotenusa del triángulo rectángulo 0, x, z) nuestro módulo al cuadrado requerido es

$$|z|^2=x^2+y^2$$

Nótese que para que esto sea una probabilidad real "normalizada", el valor de $|z|^2$ debe estar entre 0 y 1 . Esto significa que, para una amplitud adecuadamente normalizada, el punto z en el plano de Argand debe estar en alguna parte dentro del *círculo unitario* (véase fig. VI.8).

A veces, sin embargo, conviene considerar combinaciones

$$w \text{ x alternativa A} + z \text{ x alternativa B},$$

donde w y z son simplemente *proporcionales* a las amplitudes de probabilidad, y no necesitan estar dentro de este círculo. La condición para que estén *normalizadas (y* por consiguiente proporcionen las verdaderas amplitudes de probabilidad) es que la suma de los *módulos al cuadrado* sea la unidad:

$$|w|^2+|z|^2= 1$$

Si no están normalizadas de esta forma, entonces las verdaderas amplitudes para A y B son, respectivamente, $w/\sqrt{(|w|^2+|z|^2)}$ y $z/\sqrt{(|w|^2+|z|^2)}$ que *están* dentro del círculo unitario.

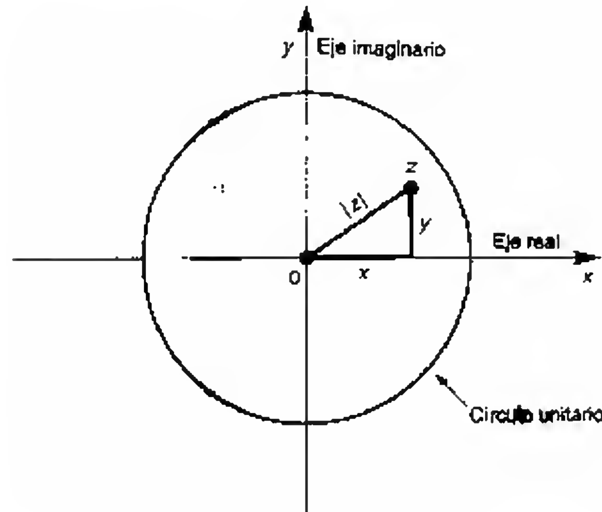


FIGURA VI.8. *Una amplitud de probabilidad representada como un punto z dentro del círculo unidad en el plano de Árgana. El cuadrado de la distancia al centro* $|z|^2$ *puede transformarse en una probabilidad real cuando los efectos se amplifican hasta el nivel clásico.*

Vemos ahora que una amplitud de probabilidad no es como una probabilidad, después de todo, sino más bien como la raíz cuadrada compleja de una probabilidad. ¿Según esto, cómo cambian las cosas cuando se amplifican los efectos del nivel cuántico hasta alcanzar el nivel clásico? Recordemos que al manipular las probabilidades y amplitudes necesitábamos multiplicarlas a veces y a veces sumarlas. El primer punto a señalar es que la operación de *multiplicación* no plantea ningún problema en nuestro paso del nivel cuántico al clásico. Esto se debe al notable hecho matemático de que el módulo al cuadrado del producto de dos números complejos es igual al producto de sus módulos al cuadrado individuales:

$$|zw|^2 = |z|^2|w|^2.$$

(Esta propiedad se sigue inmediatamente de la descripción geométrica del producto de un par de números complejos, dada en el capítulo III; pero en términos de las partes real e imaginaria $z = x + iy$, $w = u + iv$, es un pequeño milagro. Inténtelo.)

Este hecho tiene como consecuencia que si existe solamente una ruta para la partícula, por ejemplo, cuando hay sólo una rendija abierta (digamos la *s)* en el experimento de las dos rendijas, entonces podemos razonar "clásicamente", y las probabilidades vienen a ser las mismas ya se haga o no una detección adicional de la partícula en un punto intermedio (en *s*).*

Podemos tomar el módulo al cuadrado en ambas etapas o sólo al final por ejemplo

$$|A(f,s)|^2 \times |A(s,p)|^2 = |A(f,s) \times A(s,p)|^2,$$

y la respuesta para la probabilidad resultante viene a ser la misma de ambos modos.

* Esta detección debe hacerse de tal forma que no perturbe el paso de la partícula a través de *s*. Lo que podría conseguirse colocando detectores en *otros* lugares que no sean *s* e *infiriendo* el paso de la panícula por *s* cuando estos otros detectores *no* hacen click.

No obstante, si hay más de una ruta disponible (por ejemplo, si ambas rendijas están abiertas) entonces necesitamos formar una *suma* y es entonces cuando comienzan a emerger las características distintivas de la mecánica cuántica. Cuando formamos el módulo al cuadrado de la suma $w + z$, de dos números complejos w y z, normalmente no obtenemos la suma de sus módulos al cuadrado por separado; hay un "término correctivo adicional":

$$|w + z|^2 + |w|^2 + | z |^2 + 2 | w | | z | cos\theta .$$

Aquí, θ es el ángulo que el par de puntos z y w subtienden en el origen en el plano de Argand (véase fig. VI.9).

Recuérdese que el coseno de un ángulo es el cociente "cateto adyacente / hipotenusa" en un triángulo rectángulo. El lector curioso que no esté familiarizado con la fórmula anterior, puede entretenerse en deducirla directamente utilizando la geometría introducida en el capítulo III. De hecho, esta fórmula no es otra que la familiar "regla del coseno" ligeramente modificada. Es este término correctivo $2|w||z|cos\theta$. el que proporciona la *interferencia cuántica* entre opciones mecánico-cuánticas. El valor de cosθ .puede ir desde -1 a 1. Cuando θ = 0° tenemos cos θ = 1 y las dos opciones se refuerzan mutuamente, de modo que la probabilidad total es mayor que la suma de las probabilidades individuales. Cuando θ = 180° tenemos cos θ = -1 y las dos opciones tienden a cancelarse, dando una probabilidad total menor que la suma de las probabilidades individuales (interferencia destructiva). Cuando θ = 90° tenemos cos θ = 0 y estamos en una situación intermedia en la que las dos probabilidades se suman.

Para sistemas grandes o complicados, los términos correctivos generalmente se "cancelan en promedio" —debido a que el valor "promedio" de cos θ es cero— quedándonos con las reglas ordinarias de la probabilidad clásica. No obstante, en el nivel cuántico estos términos producen efectos importantes de interferencia.

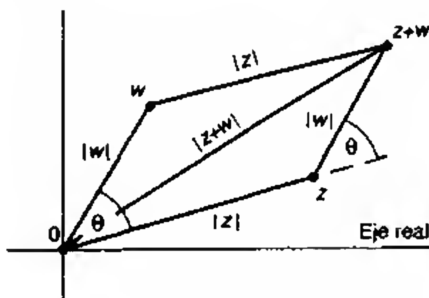


FIGURA VI.9. *Geometría relativa al término correctivo* $2|w||z|cos\theta$ *del módulo al cuadrado de la suma de dos amplitudes.*

Consideremos el experimento de la doble rendija cuando las dos rendijas estén abiertas. La amplitud para que el fotón llegue a p es una suma, $w + z$, donde

$$w = A(f,s) \times A(s,p) \qquad z = A(f,i) \times A(i,p).$$

En los puntos *más brillantes* de la pantalla tenemos $w = z$ (de modo que cos θ = 1) de donde

$$|w + z|^2 + |2w|^2 = 4|w|^2$$

que es el cuádruple de la probabilidad $|w|^2$ cuando sólo está abierta la rendija superior —y, por consiguiente, el *cuádruple* de la intensidad cuando hay un gran número de fotones, en concordancia con nuestras observaciones.

En los puntos oscuros de la pantalla tenemos $w = -z$ (de modo que cos θ = -1), de donde

$$|w + z|^2 + |w - w|^2 = 0,$$

es decir, *cero* —interferencia destructiva—, otra vez en concordancia con la observación.

En los puntos exactamente intermedios, tenemos $w = iz$ *o* $w = -iz$ (de modo que cos θ = 0), de donde

$$|w + z|^2 = |w \pm iw|^2 = |w|^2 + |w|^2 = 2|w|^2$$

que da el *doble* de la intensidad que para una sola rendija (como sería el caso para las partículas clásicas).

Al final de la próxima sección veremos cómo calcular dónde están realmente las zonas brillantes, oscuras e intermedias.

Deberíamos señalar un último punto. Cuando ambas rendijas están abiertas, la amplitud de que la partícula llegue a *p* vía *s* es en realidad $w = A(f,s) \times A(s,p)$, pero no podemos interpretar su módulo al cuadrado $|w|^2$ como la probabilidad de que la partícula atraviese "realmente" la rendija superior para llegar a *p*. Esto nos daría respuestas absurdas, especialmente si *p* está en una zona oscura de la pantalla. Pero si decidimos "detectar" la presencia del fotón en *s*, amplificando el efecto de su presencia (o ausencia) *ahí* hasta el nivel clásico, entonces *podemos* utilizar $|A(f,s)|^2$ para la probabilidad de que el fotón esté realmente presente en *s*. Tal detección, sin embargo, borraría la figura ondulatoria.

Para que tenga lugar la interferencia debemos asegurar que el paso del fotón a través de las rendijas *quede en el nivel cuántico,* de modo que *ambas* rutas opcionales deben y pueden a veces cancelarse mutuamente. En el nivel cuántico las rutas opcionales individuales tienen sólo amplitudes, no probabilidades.

EL ESTADO CUÁNTICO DE UNA PARTICULA

¿Qué tipo de imagen de la "realidad física" se nos presenta en el nivel cuántico? ¿En dónde deben poder coexistir siempre las "diferentes posibilidades" opcionales abiertas a un sistema, sumadas y ponderadas con estos extraños pesos estadísticos complejos?

Muchos físicos no confían en encontrar tal imagen. En lugar de ello, aseguran contentarse con la idea de que la teoría cuántica proporciona simplemente un procedimiento de cálculo para obtener las probabilidades y no una imagen objetiva del mundo físico. Algunos, de hecho, aseguran que de acuerdo con la teoría cuántica no es posible ni una imagen objetiva, al menos ninguna que sea consistente con los hechos físicos. Por mi parte, considero bastante injustificado tal pesimismo. En cualquier caso, sobre la base de lo que hemos discutido hasta ahora, sería prematuro adoptar esta postura. Más adelante tendremos que abordar algunas de las implicaciones más sorprendentemente enigmáticas de los efectos cuánticos, y tal vez entonces empezaremos a apreciar las razones de esa desesperación. Pero por ahora procedamos de forma más optimista y estudiemos la imagen con la que la teoría cuántica nos dice que debemos enfrentarnos.

Esta imagen es la que presenta un *estado cuántico*. Tratemos de pensar en una sola partícula cuántica. Clásicamente, una partícula está determinada por su posición en el espacio, y también

necesitamos conocer su velocidad (o, de forma equivalente, su momento) para saber qué va a hacer a continuación.

En mecánica cuántica *cada posición simple* que la partícula pudiera tener es una "alternativa" disponible. Hemos visto que todas las opciones deben combinarse de algún modo, con pesos estadísticos complejos. Esta colección de pesos estadísticos complejos describe el estado cuántico de la partícula. Es una práctica común, en teoría cuántica, utilizar la letra griega ψ ("psi") para señalar esta colección de pesos estadísticos considerada como una función compleja de la posición, llamada la *función de onda* de la partícula.

Para cada posición x, esta función de onda tiene un valor específico, denotado por $\psi(x)$, que es la amplitud de que la partícula esté en x. Podemos utilizar simplemente la letra ψ para etiquetar el estado cuántico como un todo. Estoy asumiendo la idea de que la *realidad física* de la localización de la partícula es, en realidad, su estado cuántico ψ.

¿Cómo representar la función compleja ψ? Esto es un poco difícil para todo el espacio tridimensional a un tiempo, de modo que simplificaremos un poco y supondremos que la partícula tendrá como límite situarse en una línea unidimensional, a lo largo del eje x, pongamos por caso, de un sistema de coordenadas estándar (cartesiano). Si ψ hubiera sido una función real, entonces podríamos haber imaginado un "eje *y*" perpendicular al eje x y representado la gráfica de ψ (fig. VI. 10a). Sin embargo, aquí necesitamos un "eje *y* complejo" —que sería un plano de Argand— para describir el valor de la función *compleja* ψ. Para ello, podemos utilizar, en nuestra imaginación, dos dimensiones espaciales más: por ejemplo, la dirección -y en el espacio como el eje *real* de este plano de Argand y la dirección -z como el eje *imaginario.*

Para una imagen precisa de la función de onda, representamos $\psi(x)$ como un punto en el plano de Argand —esto es, en el plano (y,z) a través de cada posición en el eje x—. A medida que x *varía,* este punto varía también y su curso describe una curva en el espacio que se enrosca en la vecindad del eje x (véase fig. VI.10b). Llamemos a esta curva la *curva* ψ de la partícula. La probabilidad de encontrar la partícula en un punto concreto x se obtiene, si hubiera un detector de partículas situado en ese punto, tomando el módulo al cuadrado de la amplitud $\psi(x)$,

$$|\psi(x)|^2$$

que es el cuadrado de la distancia de la curva ψ *al* eje x*.

* Aquí surge una dificultad técnica debido a que la probabilidad real de encontrar una partícula en un punto preciso sería nula. En lugar de ello, consideraremos que $|\psi(x)|^2$ define una *densidad de probabilidad,* lo que significa que lo que está definido es la probabilidad de encontrar una partícula dentro de algún intervalo de tamaño finito centrado en el punto en cuestión. Por lo tanto, ψ (x) define una *densidad de amplitud* y no ya una amplitud.

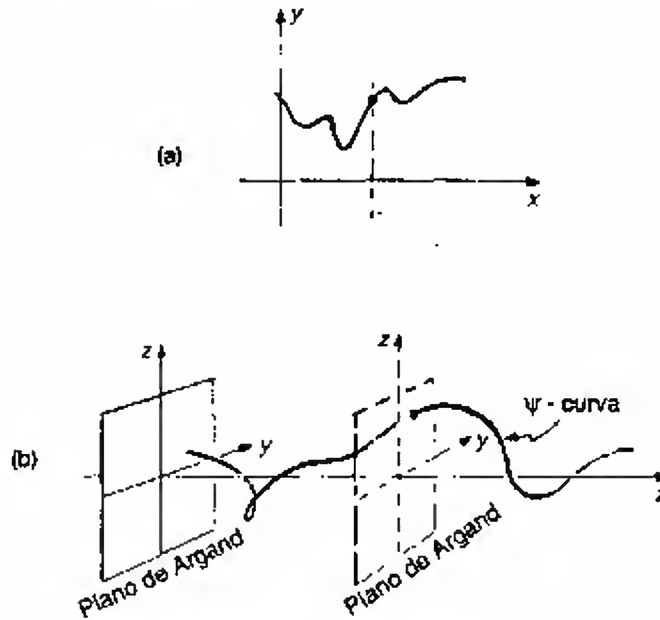


FIGURA VI. 10. (a) *Gráfica de una función real de una variable real x.* (b) *Gráfica de una función compleja y de una variable real x.*

Para formar una imagen completa de este tipo para la función de onda en todo el espacio físico tridimensional, serían necesarias *cinco* dimensiones: tres dimensiones para el espacio físico más otras dos para el plano de Argand en cada punto en el cual representar ψ *(x).* Sin embargo, nuestra imagen simplificada es aún útil. Si decidimos examinar el comportamiento de la función de onda a lo largo de cualquier línea particular en el espacio físico, podemos hacerlo tomando nuestro eje *x a* lo largo de esta línea y utilizando provisionalmente las otras dos direcciones espaciales para proporcionar los planos de Argand requeridos. Esto será útil para entender el experimento de la doble rendija.

Como mencioné antes, en física clásica necesitamos conocer la velocidad (o el momento) de una partícula para determinar lo que hará a continuación. La mecánica cuántica nos permite una notable economía: la función de onda ψ contiene *ya* las diversas amplitudes de los diferentes momentos posibles. (Algunos lectores contrariados pueden estar pensando que "ya es hora" de un poco de economía, considerando lo mucho que hemos tenido que complicar la simple imagen clásica de una partícula puntual. Aunque tengo mucha simpatía por esos lectores, sugiero que aprovechen los bocados que se les ofrecen, pues lo peor está por llegar.) ¿Cómo es que las amplitudes de las velocidades están determinadas por ψ ? En realidad, es mejor pensar en las amplitudes de los *momentos.* (Recordemos que el momento es la velocidad multiplicada por la masa de la partícula). Lo que se hace es aplicar el llamado *análisis armónico* a la función. Estaría fuera de lugar aquí explicar eso en detalle, si no estuviera estrechamente relacionado con lo que se hace con los sonidos musicales.

Una onda de forma cualquiera puede descomponerse como una suma de diferentes "armonías" (de ahí el término "análisis armónico") que son los "tonos puros" de las distintas notas (es decir, distintas frecuencias puras). En el caso de la función de onda ψ , los "tonos puros" corresponden a los diferentes valores posibles del momento que pudiera tener la partícula, y el tamaño de la

contribución a ψ de cada "tono puro" proporciona la amplitud de ese valor del momento. Los mismos "tonos puros" se denominan *estados de momento.*

¿Qué apariencia tiene un estado de momento como curva ψ *?* Tiene el aspecto de un *sacacorchos,* para el que el nombre matemático oficial es *hélice* (fig. VI.11).* Aquellos sacacorchos que están enroscados apretadamente corresponden a momentos grandes, y los que apenas se enroscan dan momentos muy pequeños. Hay un caso límite en el que no se enroscan y la curva y es una línea recta: el caso de momento cero. La famosa *relación de Planck* está implícita en esto. Un paso de rosca pequeño significa corta longitud de onda y alta *frecuencia* y, por lo tanto, alto momento y alta *energía;* y un paso de rosca grande significa baja frecuencia y baja energía, siendo la energía E siempre proporcional a la frecuencia v $(E = hv)$. Si los planos de Argand están orientados de la forma normal (es decir, con la descripción x, y, z dada arriba, con los usuales ejes dextrógiros), entonces los momentos que apuntan en la dirección positiva del eje x corresponden a los sacacorchos dextrógiros (que son los sacacorchos de tipo usual).

A veces es útil describir los estados cuánticos no en términos de funciones de onda ordinarias, como se hizo arriba, sino en términos de funciones de onda del *momento.* Esto equivale a considerar la descomposición de ψ en términos de los diversos estados de momento y a construir una nueva función Ψ, que esta vez es una función del momento p en lugar de

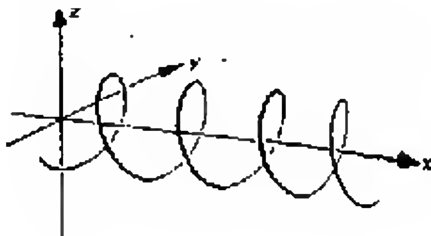


FIGURA VI.11. *Un estado de momento tiene una curva* ψ *que es un sacacorchos.*

la posición x, cuyo valor $\Psi(p)$, para cada p, da la medida de la contribución del estado de momento p a ψ. (El espacio de los p se llama *espacio de momentos.)* La interpretación de Ψ es que, para cada particular elección de p, el número complejo $\Psi(p)$ da la *amplitud de que la partícula tenga un momento p.*

En matemáticas, a la función $\Psi(p)$ se le llama la *transformada de Fourier* de ψ y viceversa, por el ingeniero y matemático francés Joseph Fourier (1768-1830); sólo haré aquí algunos comentarios sobre esta relación.

El primer punto es que existe una notable simetría entre ψ y Ψ. Para regresar a ψ desde Ψ, aplicamos efectivamente el mismo procedimiento que aplicamos para obtener Ψ a partir de ψ. Ahora es Ψ la que se somete a análisis armónico, Los "tonos puros" (es decir, los *sacacorchos* en la imagen del espacio de momentos) se llaman ahora *estados de posición.* Cada posición x determina dicho "tono puro" en el espacio de momentos, y el tamaño de la contribución a Ψ de este "tono puro" da el valor $\psi(x)$.

* En términos de una descripción analítica más estándar, cada uno de nuestros *sacacorchos* (es decir, estados de momento) vendría dado por una expresión $\psi = e^{ipx/\hbar} = \cos(ipx/\hbar) + i\operatorname{sen}(ipx/\hbar)$ *(véase* capítulo III, donde p es el valor del momento en cuestión).

Un estado de posición corresponde, en la imagen ordinaria del espacio de posiciones, a una función ψ que tiene un pico muy agudo en el valor de *x* en cuestión, siendo nulas las amplitudes para otros valores diferentes del *x* dado. Semejante función se denomina *función delta* (de Dirac), aunque técnicamente no es una "función" en el sentido ordinario, ya que toma un valor infinito en *x.* Análogamente, los estados de momento *(sacacorchos* en la imagen del espacio de posiciones) dan funciones delta en la imagen del espacio de momentos (véase fig. VI. 12) Vemos así que la transformada de Fourier de un sacacorchos es una función delta, y *viceversa.*

La descripción en el espacio de posiciones es útil cuando intentamos realizar mediciones de la posición de la partícula, lo que equivale a hacer algo que amplifica los efectos de las diferentes posiciones posibles de la partícula hasta el nivel clásico. En términos generales, las fotocélulas y las placas fotográficas realizan mediciones de la posición para los fotones. Asimismo, la descripción en el espacio de momentos es útil cuando nos proponemos medir el momento de la partícula, es decir, amplificar los efectos de los diferentes momentos posibles hasta el nivel clásico. (Pueden utilizarse los efectos de retroceso o de difracción en cristales para medidas de momento.) En cada caso, el módulo al cuadrado de la función de onda correspondiente (ψ o Ψ) da la probabilidad requerida para el resultado de la medida también requerida.

Terminaremos esta sección volviendo una vez más al experimento de la doble rendija.

Hemos aprendido que, según la mecánica cuántica, incluso una simple partícula debe comportarse por sí misma como toda una onda. Esta onda viene descrita por la función de onda ψ . Las ondas más uniformemente onduladas son los estados de momento. En el experimento de la doble rendija suponíamos fotones de una frecuencia definida. De ese modo, la función de onda del fotón se compone de estados de momento en diferentes direcciones, en los que el paso de rosca es el mismo para todos los sacacorchos, y esa distancia es la *longitud de onda.* (La longitud de onda está determinada por la frecuencia.)

Cada función de onda de un fotón se extiende inicialmente desde la fuente *f.* Y, si no se hace ninguna detección en las rendijas, atraviesa a ambas en su camino hasta la pantalla. Pero sólo una pequeña parte de esta función de onda emerge de las rendijas, y consideramos que cada rendija actúa como una nueva fuente a partir de la cual se extienden funciones de onda independientes. Estas dos porciones de la función de onda interfieren entre sí. De modo que, cuando alcanzan la pantalla, hay lugares en donde las dos porciones se suman y lugares en donde se cancelan. Para descubrir en dónde se suman las ondas y en dónde se cancelan, tomemos algún punto *p* de la pantalla y examinemos las líneas rectas que van a *p* desde cada una de las dos rendijas *s* e *i.*

A lo largo de la línea *sp* tenemos un *sacacorchos,* y a lo largo de la línea *ip* tenemos otro. (Los tenemos también a lo largo de las líneas *fs* y *fi,* pero si suponemos que la fuente está a la misma distancia de las dos rendijas, entonces *en* las rendijas los sacacorchos habrán girado la misma cantidad.)

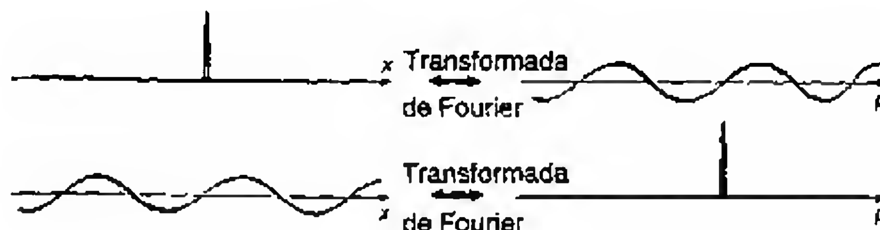


FIGURA VI. 12. *Las funciones delta en el espacio de posiciones corresponden a , sacacorchos en el espacio de momentos, y viceversa.*

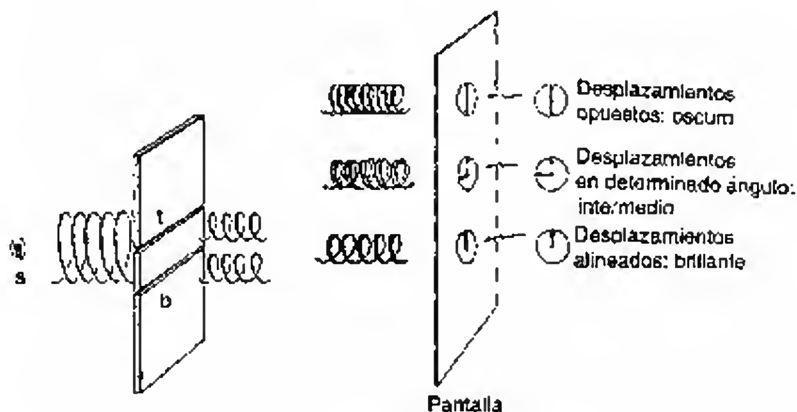


FIGURA VI. 13. *El experimento de la doble rendija, analizado en términos de la descripción de los estados de momento del fotón como un* sacacorchos.

Ahora, las cantidades que han girado los sacacorchos cuando alcanzan la pantalla en *p* dependerán de las longitudes de las líneas *sp* e *ip*. Cuando estas longitudes difieran en un número entero de longitudes de onda —entonces, en *p*—, ambos sacacorchos tendrán el *mismo* desplazamiento respecto a sus ejes (esto es, $\theta = 0°$, dónde θ es el mismo que en la sección previa), de modo que las amplitudes respectivas se sumarán y tendremos un punto *brillante*.

Cuando estas longitudes difieran en un número entero de longitudes de onda más media longitud de onda —en *p*—, los desplazamientos de ambos sacacorchos respecto a sus ejes serán *opuestos* ($\theta = 180°$), con lo que las amplitudes respectivas se cancelarán y obtendremos un punto *oscuro*. En todos los demás casos habrá algún ángulo entre los desplazamientos de los sacacorchos cuando alcancen *p*, de modo que las amplitudes se suman de una forma intermedia y obtenemos una región de intensidad intermedia (véase fig. VI. 13.)

EL PRINCIPIO DE INCERTIDUMBRE

La mayoría de los lectores habrá oído hablar del *principio de incertidumbre* de Heisenberg, según el cual no es posible medir (es decir, magnificar en el nivel clásico) al mismo tiempo, con precisión la posición y el momento de una partícula. Peor aún: existe un *límite absoluto* para el producto de estas precisiones, llamémoslas Δx y Δp, respectivamente, que viene dado por la relación

$$\Delta x \, \Delta p \geq \hbar$$

Esta fórmula nos dice que cuanto más precisa sea medida la posición *x,* con menor precisión se puede determinar el momento *p,* y *viceversa.* Si la posición fuera medida con precisión *infinita,* entonces el momento quedaría *totalmente* indeterminado. Por el contrario, si el momento se mide exactamente, entonces la localización de la partícula queda totalmente indeterminada.

Para hacernos una idea del tamaño del límite dado por la relación de Heisenberg, supongamos que la posición de un electrón se ha medido con la precisión de un manómetro (10^{-9} m). Entonces, el momento se haría tan indeterminado que no podríamos esperar que, un segundo después, el electrón estuviera a menos de 100 kilómetros de distancia.

En algunas descripciones, nos llevan a creer que esto se debe simplemente a una torpeza inherente al proceso de medida. Consiguientemente, en el caso del electrón recién considerado, el intento para localizarlo le da — según este modo de verlo — un "empujón" aleatorio de tal intensidad que el electrón sale probablemente despedido a una gran velocidad (en el orden de magnitud indicado por el principio de Heisenberg). En otras descripciones se nos dice que la incertidumbre es una propiedad de la propia partícula y que su movimiento tiene una aleatoriedad inherente, lo que significa que su comportamiento es intrínsecamente impredecible en el nivel cuántico. Todavía hay otras exposiciones en las que se nos informa que una partícula cuántica es algo incomprensible, a causa de lo cual los mismos conceptos de posición clásica y momento clásico son inaplicables. Yo no me siento a gusto con ninguna de ellas. La primera es algo confusa, la segunda es ciertamente errónea, y la tercera, indebidamente pesimista.

¿Qué nos dice en realidad la descripción de la función de onda? Recordemos primero nuestra descripción de un estado de momento. Este es el caso en que el momento está especificado exactamente. La curva ψ es un sacacorchos que permanece enroscado siempre a la misma distancia del eje. Por consiguiente, las amplitudes para los diferentes valores de la posición tienen todas los mismos módulos al cuadrado. Así, si se realiza una medida de la posición, la probabilidad de encontrar la partícula en un punto es la misma que la de encontrarla en otro cualquiera. La posición de la partícula está completamente indeterminada.



FIGURA VI. 14. *Paquetes de ondas. Están localizados tanto en el espacio de posiciones como en el de momentos.*

¿Qué hay sobre el estado de posición? En este caso, la curva ψ es una función delta. La partícula está localizada exactamente —o sea en la posición del pico de la función delta— y son nulas las amplitudes para las otras posiciones. Las amplitudes de momento se obtienen mejor si consideramos la descripción en el espacio de momentos, donde ahora la curva Ψ es un sacacorchos y donde también ahora las diferentes amplitudes de momento son las que tienen igual módulo al cuadrado. Al realizar una medida del momento de la partícula, el resultado estaría entonces completamente indeterminado.

Es interesante examinar un caso intermedio en el que las posiciones y los momentos están parcialmente limitados, aunque esto sólo hasta un grado consistente con la relación de Heisenberg. La curva ψ y la correspondiente curva Ψ para este caso (transformadas de Fourier una de otra) se ilustran en la fig. VI. 14. Nótese que la distancia de cada curva al eje es apreciable sólo en una región bastante pequeña. Lejos de ésta, la curva se ciñe muy estrechamente al eje. Esto significa que los módulos al cuadrado son de tamaño apreciable en una región muy limitada, tanto en el espacio de posiciones como en el espacio de momentos. Por lo tanto, la partícula puede estar algo localizada en el espacio pero hay una cierta dispersión. Análogamente, el momento está bastante definido. La partícula se mueve con una velocidad definida y la dispersión de las posibles posiciones no crece tan rápido con el tiempo. Un estado cuántico de este tipo se conoce como *paquete de ondas.* Frecuentemente se toma como la mejor aproximación en teoría cuántica a una partícula clásica, pero la dispersión en los valores del momento (es decir, de la velocidad) implica que un paquete de ondas se *dispersará* con el tiempo. Cuanto más localizada esté la posición de partida, más rápidamente se ensanchará.

LOS PROCEDIMIENTOS DE EVOLUCIÓN U Y R

En esta descripción del desarrollo temporal de un paquete de ondas está implícita *la ecuación de Schrödinger* que nos dice cómo evoluciona la función de onda con el tiempo: si descomponemos ψ en estados de momento ("tonos puros") cada uno de los componentes individuales se moverá con una velocidad que es c^2 dividida entre la velocidad de una partícula clásica que tuviera el momento en cuestión. De hecho, la ecuación matemática de Schrödinger se escribe de forma más concisa que esto. Veremos su forma exacta más adelante. Se asemeja algo a las ecuaciones de Hamilton o de Maxwell (y tiene estrechas relaciones con ambas). Y, como aquellas ecuaciones, da una evolución *completamente determinista* de la función de onda una vez que esta función de onda se especifica para un instante cualquiera.

Considerando que ψ describe la "realidad" del mundo, y que está gobernada por la ecuación determinista de Schrödinger, no tenemos nada de este indeterminismo porque tal es una característica inherente a la teoría cuántica. Nos referiremos a este proceso como el proceso de evolución **U**. Sin embargo, cada vez que amplificamos los efectos cuánticos hasta el nivel clásico y "hacemos una medición", cambiamos las reglas. Ahora *no* utilizamos **U** sino que en su lugar adoptamos un procedimiento completamente diferente, que llamaré **R**, el cual consiste en formar los cuadrados de los módulos de las amplitudes cuánticas para obtener probabilidades clásicas.[4] Es el procedimiento **R**, y *sólo* **R**, el que introduce incertidumbres y probabilidades en la teoría cuántica.

El proceso determinista **U** parece ser la parte de la teoría cuántica de mayor interés para el trabajo de los físicos. No obstante, los filósofos están más intrigados por la reducción **R** del vector de estado (o, como a veces se le describe gráficamente: *colapso de la función de onda)* no

[4] Estos dos procedimientos de evolución se describieron en una obra clásica del famoso matemático húngaro-estadounidense John von Neumann (1955). Su "proceso 1" es el que yo he llamado **R** —"reducción del vector de estado"— y su "proceso 2" es **U** —"evolución unitaria"— (lo que significa que las amplitudes de probabilidad se conservan efectivamente en la evolución). De hecho, existen otras —aunque equivalentes— descripciones de la evolución **U** del estado cuántico, en las que se podría no utilizar el término "ecuación de Schrödinger". En la *imagen de Heisenberg,* por ejemplo, el estado se describe de modo que parece no evolucionar en absoluto, siendo considerada la evolución dinámica como un continuo desplazamiento de los significados de las coordenadas de posición-momento. Las diversas distinciones no nos interesan aquí pues las diferentes descripciones del proceso **U** son completamente equivalentes.

determinista. Ya sea que consideremos **R** simplemente como un cambio en el "conocimiento" disponible del sistema, o ya sea que lo tomemos (como yo lo hago) como algo "real", disponemos de dos modos matemáticos completamente *distintos* para describir cómo con el tiempo cambia el vector de estado de un sistema físico.

En efecto, **U** es completamente determinista, mientras que **R** es una ley probabilista. **U** mantiene la superposición compleja cuántica pero **R** la viola totalmente. **U** actúa de una forma continua, pero **R** es descaradamente discontinuo. Según los procedimientos estándar de la mecánica cuántica, es imposible deducir **R** como un ejemplo complicado de **U** estamos ante un procedimiento *diferente* de **U** que proporciona la otra "mitad" de la interpretación del formalismo cuántico. Todo el no determinismo de la teoría procede de **R** y no de **U**. Ambos, **U** y **R**, son necesarios para todos los acuerdos entre la teoría cuántica y los hechos que se desprenden de la observación.

Volvamos a nuestra función de onda ψ . Supongamos que es un estado de momento. Seguirá siendo ese mismo estado de momento durante el resto del tiempo, en tanto la partícula no tenga interacciones con algo. (Esto es lo que nos dice la ecuación de Schrödinger.) En el instante en que decidamos "medir su momento" seguiremos teniendo la misma respuesta definida. Aquí no hay probabilidades. La predecibilidad es tan estricta como en la teoría clásica, pero supongamos que en alguna etapa decidimos medir (es decir, amplificar hasta el nivel clásico) la posición de la partícula: se nos presentará una colección de amplitudes de probabilidad cuyos módulos deben ser elevados al cuadrado. En este punto dominan las probabilidades y la incertidumbre sobre el resultado que va a dar la medida. Tal incertidumbre va de acuerdo con el principio de Heisenberg.

Supongamos, por el contrario, que ψ parte de un estado de posición (o de muy cerca de un estado de posición). La ecuación de Schrödinger nos dice que *no* permanecerá en un estado de posición, sino que se dispersará rápidamente. De todas formas, el *modo* en que se dispersa está completamente fijado por esta ecuación. No hay nada indeterminado o probabilístico en este comportamiento.

En principio, existen experimentos que podrían verificar este hecho (abundaremos en esto adelante), pero si imprudentemente decidimos medir el momento, entonces encontraremos amplitudes para todos los diferentes valores posibles del momento que tienen iguales módulos al cuadrado y habrá incertidumbre completa acerca del resultado del experimento, de nuevo en acuerdo con el principio de Heisenberg.

De la misma manera, si ψ parte como un paquete de ondas, su evolución está determinada por la ecuación de Schrödinger y, en principio, se Podrían idear experimentos para seguirle la pista a ese hecho. Pero cuando decidimos medir la partícula de una forma *diferente* —por ejemplo, medir su posición o momento—, encontramos que se introducen las incertidumbres, de nuevo en acuerdo con el principio de Heisenberg, con probabilidades dadas por los cuadrados de los módulos de las amplitudes. Todo esto puede parecer extraño o misterioso, pero no incomprensible. Hay mucho en esta imagen que está gobernado por leyes claras y precisas. No hay, en cambio, una idea clara sobre cuándo deberíamos invocar la regla probabilista **R** en lugar de la determinista **U**.

¿En qué consiste el "hacer una medición"? ¿Por qué (y cuándo) los cuadrados de los módulos de las amplitudes se "convierten en probabilidades"? ¿Puede entenderse el "nivel clásico" desde el

punto de vista de la mecánica cuántica? Estas son preguntas profundas y enigmáticas que serán abordadas en este capítulo.

¿PARTÍCULAS EN DOS LUGARES A LA VEZ?

En las descripciones anteriores he adoptado una visión bastante más "realista" de la función de onda, que la que quizá es común entre los físicos cuánticos. He adoptado el punto de vista de que el estado "objetivamente real" de una partícula viene descrito por su función de onda ψ . Parece que mucha gente encuentra que ésta es una postura a la que es difícil adherirse con seriedad. Un motivo para ello parece radicar en que tal postura implica considerar que las partículas individuales tienen una extensión espacial, en lugar de que están concentradas en puntos simples.

Para un estado de momento, esta dispersión llega a su punto más extremo porque ψ está uniformemente distribuida sobre la totalidad del espacio. Más que pensar que la propia partícula se extiende en el espacio, la gente prefiere considerar que su posición está "completamente indeterminada", de modo que tan probable es que la partícula esté en un lugar como que esté en otro.

Pero ya hemos visto que la función de onda no proporciona simplemente una distribución de probabilidad para las distintas posiciones, sino toda una distribución de *amplitudes* para las diferentes posiciones. Si conocemos esta distribución de amplitudes (es decir, la función ψ), entonces conocemos —a partir de la ecuación de Schrödinger— la forma precisa en que evolucionará el estado de la partícula. Necesitamos esta idea de partícula "extendida" para que su "movimiento" (esto es, la evolución de ψ en el tiempo) esté determinado. Y si *aceptamos* esta idea, veremos que el movimiento de la partícula *está* determinado de manera precisa. La "concepción probabilista" con respecto a $\psi(x)$ se haría apropiada si realizáramos una medida de la posición de la partícula y $\psi(x)$ se utilizara *entonces* sólo como un módulo al cuadrado: $|\psi(x)|^2$

Parece que tenemos que aceptar esta imagen de una partícula que puede extenderse sobre grandes regiones del espacio y permanecer extendida hasta que se lleve a cabo la siguiente medida de posición.

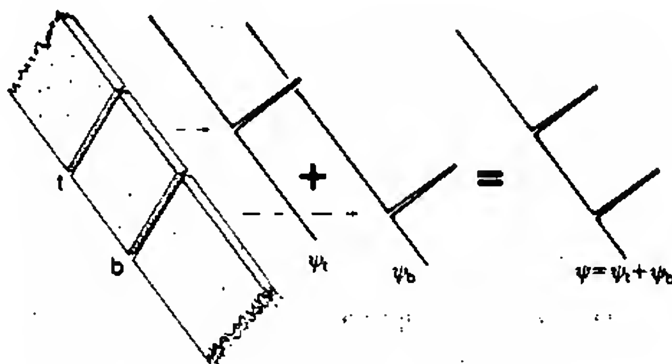


FIGURA VI. 15. *La función de onda del fotón tiene picos en dos lugares a la vez cuando emerge del par de rendijas.*

Incluso cuando está localizada como un estado de posición, una partícula empieza a dispersarse en el instante siguiente. Un estado de momento puede parecer difícil de aceptar como una imagen de la "realidad" de la existencia de la partícula, pero es todavía más difícil aceptar como

"real" el estado con *doble pico* que ocurre cuando la partícula emerge inmediatamente después de atravesar un par de rendijas (fig. VI. 15). En la dirección vertical, la forma de la función de onda y tendría un pico agudo en cada una de las rendijas, y es la suma* de una función de onda ψ_s, que tiene un pico en la rendija superior, y ψ_i, que tiene un pico en la rendija inferior:

$$\psi(x) = \psi_s(x) + \psi_i(x).$$

Si tomamos y representando la "realidad" del estado de la partícula, entonces debemos aceptar que la partícula "está" de hecho en *dos* lugares a la vez. Según este enfoque la partícula *ha atravesado ambas rendijas a la vez.*

No hay que olvidar la objeción generalizada a la idea de que la partícula "atraviesa ambas rendijas a la vez". Si realizamos una medida en las rendijas para determinar cuál de las rendijas ha atravesado, encontraremos siempre que la partícula *entera* está en una u otra de las rendijas. Pero esto aparece así debido a que estamos realizando una *medida de posición* sobre la partícula y, por lo tanto, ψ proporciona *ahora* simplemente una distribución de probabilidades $|\psi|^2$ para la posición de la partícula de acuerdo con el procedimiento del módulo cuadrado, por lo cual hallaremos que está sólo en un lugar u otro. Pero existen varios tipos de medición que *podríamos* realizar en las rendijas, *además* de las medidas de posición. Para aquellas, necesitaríamos conocer la función de onda *y* con sus dos picos, y no sólo $|\psi|^2$, para diferentes posiciones *x.* Semejante medida podría distinguir el estado con dos picos

$$\psi = \psi_s + \psi_i .$$

dado arriba, de otros estados de dos picos, tales como

$$\psi_s - \psi_i .$$

o

$$\psi_s + i\psi_i .$$

(Véase fig. VI. 16. para las curvas ψ en cada uno de estos diferentes casos.) Puesto que existen medidas que distinguen entre diversas posibilidades, todas éstas pueden ser asimismo *diferentes* modos "realmente" posibles de la existencia de la partícula.

Las rendijas no tienen que estar próximas entre sí para que un fotón las atraviese "ambas al mismo tiempo". Para comprobar que una partícula cuántica puede estar en "dos lugares a la vez", no importa cuál sea la distancia entre éstos, consideremos un montaje un poco diferente del experimento de la doble rendija.

Como antes, tenemos una lámpara que emite luz monocromática de fotón en fotón, pero en lugar de dejar que la luz atraviese el par de rendijas la reflejamos en un espejo semirreflectante que forma un ángulo de 45° con la dirección del haz luminoso. (Un espejo semirreflectante es un espejo que refleja exactamente la mitad de la luz que incide sobre él, mientras que la mitad restante se transmite a través del espejo.)

Después de su encuentro con el espejo, la función de onda del fotón se descompone en dos, con una parte reflejada lateralmente y la otra que continúa en la misma dirección de partida del fotón.

---

* La descripción mecano-cuántica más usual dividiría esta suma entre un factor de normalización —en este caso $\sqrt{2}$, para dar $(\psi_s + \psi_i)/\sqrt{2}$ —, pero por ahora no hay necesidad de ello.

Como en el caso del fotón emergiendo de las dos rendijas, la función de onda tiene otra vez dos picos, pero ahora mucho más separados, un pico que describe el fotón reflejado y otro pico que describe el fotón transmitido. (Véase fig VI. 17.) Además, a medida que pasa el tiempo, la separación entre los picos se hace mayor, incrementándose sin límite.

Imaginemos que las dos porciones de la función de onda escapan al espacio exterior y que esperamos durante todo un año. Entonces los dos picos de la función de onda del fotón estarán separados más de un año luz. De algún modo, el fotón ha estado en dos lugares a la vez, a más de un año luz de distancia.

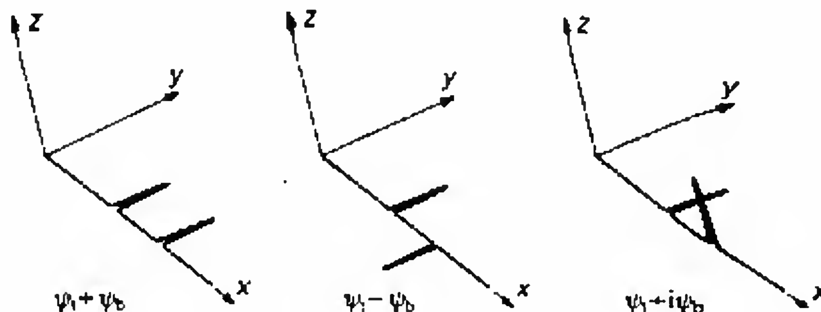


FIGURA VI. 16. *Tres formas diferentes en las que una función de onda de un fotón puede tener un doble pico.*

¿Existe alguna razón para tomar esta imagen en serio? ¿No podemos considerar simplemente que el fotón tiene 50% de probabilidades de estar en uno de los lugares y 50% de probabilidades de estar en el otro? ¡No! Por mucho que haya viajado, siempre existe la posibilidad de que las dos partes del haz sean reflejadas hacia atrás, de modo que vuelvan a encontrarse y produzcan efectos de interferencia que no podrían resultar de una ponderación probabilista de las dos posibilidades.

Supongamos que cada parte del haz encuentra un espejo totalmente reflectante, colocado en un ángulo apropiado para que los rayos se junten, y en el punto de encuentro se coloca otro espejo semirreflectante, en el mismo ángulo que el primero. En las direcciones finales de los dos haces se colocan dos fotocélulas (véase fig. VI. 18.) ¿Qué encontramos? Si fuera simplemente el caso de que hubiera 50% de probabilidades de que el fotón siguiera una ruta y 50% de probabilidades de que siguiera la otra, entonces deberíamos encontrar 50% de probabilidades de que uno de los detectores registre el fotón y 50% de probabilidades de que lo haga el otro. Sin embargo, no es esto lo que sucede.

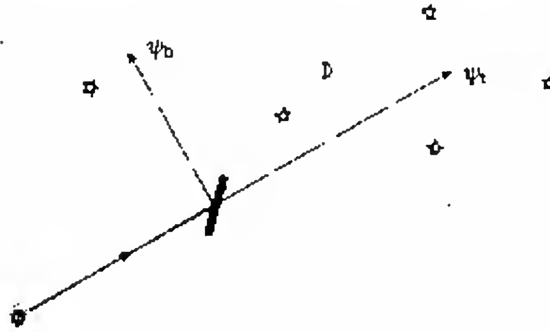


FIGURA VI. 17. *Los dos picos de una función de onda podrían estar a años luz de distancia. Esto puede conseguirse por medio de un espejo semirreflectante.*

Si las dos rutas tienen la misma longitud entonces resulta que hay 100% de probabilidades de que el fotón llegue al detector A, colocado en la dirección del movimiento inicial del fotón, y 0% de probabilidades de que llegue al otro detector B. Entonces, el fotón incidirá con *certeza* en el detector A. (Podemos verlo utilizando la descripción del sacacorcho, como en el experimento de la doble rendija.)

Por supuesto, nunca se han llevado a cabo experimentos con caminos de una longitud cercana a un año luz, pero nadie (entre los físicos cuánticos convencionales) pone en duda el resultado establecido. En realidad, los experimentos se han llevado por caminos de una longitud de varios metros, y sus resultados están de completo acuerdo con las predicciones mecánico-cuánticas *(cfr.* Wheeler, 1983).

¿Qué nos dice esto acerca de la *realidad* del estado de existencia del fotón entre su primer y su último encuentro con el espejo semirreflectante? Parece inevitable aceptar que, en algún sentido, el fotón ha *recorrido* ambas rutas al mismo tiempo.

En efecto, si se colocara una pantalla absorbente en cualquiera de las dos rutas, se haría igualmente probable alcanzar A o B, pero cuando ambas rutas están abiertas (y tienen la misma longitud) sólo puede alcanzarse A. El bloqueo de una de las rutas hace realmente posible que se alcance B. Con ambas rutas abiertas, el fotón "sabe" de alguna manera que *no* está permitido llegar a B, así que efectivamente debe tener alguna idea sobre ambas rutas.

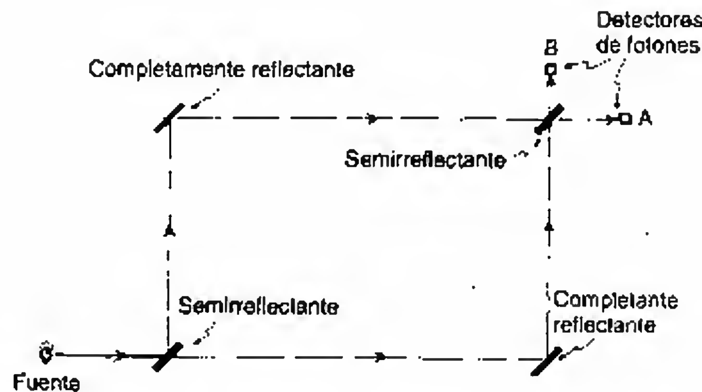


FIGURA VI. 18. *Los dos picos de una función con doble pico no pueden ser considerados simplemente como probabilidades ponderadas de que el fotón tenga una localización u otra. Es posible hacer que las dos rutas que toma el fotón se interfieran mutuamente.*

La concepción de Niels Bohr de que no puede asociarse ningún "significado" objetivo a la existencia del fotón entre los instantes de las medidas, me parece una idea demasiado pesimista como para adoptarla con respecto a la realidad del estado del fotón. La mecánica cuántica nos proporciona una *función de onda* para describir la "realidad" de la posición del fotón, y la función de onda del fotón —entre los espejos semirreflectantes— es simplemente un estado con dos picos cuya distancia puede ser, a veces, muy considerable.

Notamos, también, que el simple "estar en dos lugares concretos a la vez" no es una descripción completa del estado del fotón: necesitamos poder distinguir el estado $\psi_s+\psi_i$, del estado $\psi_s-\psi_i$, pongamos por caso —o de $\psi_s+i\psi_i$ —, donde ψ_s y ψ_i, se refieren a las posiciones del fotón en cada una de las dos rutas (ahora "transmitida" y "reflejada", respectivamente). Es este tipo de distinción el que determina si el fotón, cuando llega al último espejo semirreflectante, alcanzará A con certeza o alcanzará B con certeza (o alcanzará A o B con una probabilidad intermedia).

Esta característica enigmática de la realidad cuántica —es decir, que debemos tomar en serio el que, de varias formas (distintas), una partícula pueda "estar en dos lugares a la vez"— surge del hecho de que debemos ser capaces de sumar estados cuánticos utilizando pesos probabilísticos complejos, si queremos obtener otros estados cuánticos.

Este tipo de superposición de estados es una característica general —e importante— de la mecánica cuántica, conocida como *superposición lineal cuántica.* Y es la que nos permite componer estados de momento a partir de estados de posición, o estados de posición a partir de estados de momento.

En tales casos, la superposición lineal se aplica a una colección *infinita* de estados diferentes, es decir, a todos los estados de posición y a todos los estados de momento. Pero, como veremos, la superposición lineal cuántica resulta bastante enigmática cuando se aplica a sólo un *par* de estados.

La regla es que dos estados *cualesquiera,* independientemente de lo distintos que puedan ser, pueden coexistir en cualquier superposición lineal compleja.

En realidad, cualquier objeto físico, constituido él mismo por partículas individuales, debería poder existir en semejante superposición de estados ampliamente separados en el espacio y, por lo tanto, "estar en dos lugares a la vez." El formalismo de la mecánica cuántica no hace distinción, a este respecto, entre partículas simples y sistemas complicados de muchas partículas. ¿Por qué, entonces, no tenemos experiencia de que cuerpos macroscópicos, digamos bolas de cricket, o incluso personas, tengan dos localizaciones completamente diferentes al mismo tiempo?

Esta es una cuestión profunda, y la teoría cuántica actual no nos proporciona una respuesta realmente satisfactoria.

Para un objeto tan sustancial como una bola de cricket, debemos considerar que el sistema está "en el nivel clásico" o, como se suele expresar, habrá que hacer una "observación" o "medida" sobre la bola de cricket. Entonces las amplitudes de probabilidad complejas que ponderan nuestra superposición lineal deberán tener sus módulos elevados al cuadrado y ser tratadas como probabilidades que describen opciones reales. Sin embargo, esta realidad supone una petición de principio del *por qué* podemos cambiar así nuestras reglas cuánticas de **U** a **R**. Volveré más adelante a este punto.

ESPACIO DE HILBERT

Recuérdese que en el capítulo V se introdujo el concepto de *espacio de fases* para la descripción de un sistema físico. Un punto simple del espacio de fases representaba el estado (clásico) de un sistema físico entero.

En la teoría cuántica el concepto análogo apropiado es el de un *espacio de Hilbert.*[*] Ahora, un simple punto del espacio de Hilbert representa el estado *cuántico* de un sistema entero. Necesitaremos hacernos una idea de la estructura matemática de un espacio de Hilbert. Espero que el lector no se desanime por esto. No hay nada que sea muy complicado matemáticamente en lo que voy a decir, aunque algunas de las ideas pueden resultar poco familiares.

La propiedad más peculiar de un espacio de Hilbert es que constituye lo que se llama un *espacio vectorial* —en realidad, un espacio vectorial *complejo*—. Esto significa que podemos *sumar* dos elementos del espacio y obtener otro, del mismo espacio; y que también podemos realizar estas sumas con pesos estadísticos complejos. Podemos hacer esto porque estas son las operaciones de la *superposición lineal cuántica* que acabamos de considerar, a saber, las operaciones que nos dan $\psi_s+\psi_i$, $\psi_s-\psi_i$, $\psi_s+i\psi_i$, etc., para el fotón anterior. En esencia, todo lo que queremos decir al utilizar la expresión "espacio vectorial complejo" es que podemos formar sumas ponderadas de este tipo.[5]

Será conveniente adoptar una notación (debida a Dirac) que nos sirva para presentar los elementos del espacio de Hilbert —o *vectores de estado*— mediante algún símbolo en un paréntesis angular tales como $|\psi\rangle$, $|\chi\rangle$, $|\varphi\rangle$, $|1\rangle$, $|2\rangle$, $|3\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, $|\rightarrow\rangle$, $|\%\rangle$, etc. Así, estos símbolos representarán también estados cuánticos.

Para la operación de suma de dos vectores de estado escribimos

$$|\psi\rangle + |\chi\rangle$$

y ponderada con, números complejos *w* y *z:*

$$w|\psi\rangle + z|\chi\rangle$$

(donde $w|\psi\rangle$ significa $w \times |\psi\rangle$, etc.). En consecuencia, las combinaciones anteriores $\psi_s+\psi_i$, $\psi_s-\psi_i$, $\psi_s+i\psi_i$, se escriben ahora como $|\psi_s\rangle+|\psi_i\rangle$, $|\psi_s\rangle-|\psi_i\rangle$, $|\psi_s\rangle+i|\psi_i\rangle$, respectivamente. Podemos también multiplicar sencillamente un *simple* estado $|\psi\rangle$ por un número complejo *w* para obtener:

[*] David Hilbert, a quien ya hemos encontrado en capítulos anteriores, introdujo este importante concepto —para el caso de infinitas dimensiones— mucho antes del descubrimiento de la mecánica cuántica, y para un propósito matemático completamente diferente.

[5] Para ser completos deberíamos especificar también todas las leyes algebraicas requeridas que, en la notación (de Dirac) utilizada en el texto, son:

$|\psi\rangle+|\chi\rangle = |\chi\rangle+|\psi\rangle$, $\quad |\psi\rangle + (|\chi\rangle + |\varphi\rangle) = (|\psi\rangle + |\chi\rangle) + |\varphi\rangle$,

$(z+w)\,|\psi\rangle = z\,|\psi\rangle + w\,|\psi\rangle$, $\quad z(|\psi\rangle+|\chi\rangle) = z\,|\psi\rangle + z\,|\chi\rangle$,

$z\,(w\,|\psi\rangle) = (zw)\,|\psi\rangle$, $\quad 1|\psi\rangle = |\psi\rangle$,

$|\psi\rangle + 0 = |\psi\rangle$ $\quad 0|\psi\rangle = 0$, *y* $z0 = 0$

$$w\,|\psi\rangle$$

(Este es en realidad un caso particular de lo anterior, cuando $z = 0$.)

Ya antes admitimos combinaciones con pesos estadísticos complejos en las que w y z no necesitan ser las amplitudes de probabilidad reales sino que son simplemente *proporcionales* a dichas amplitudes. En consecuencia, adoptamos la regla de que podemos multiplicar globalmente un vector de estado por un número complejo distinto de cero y el estado físico sigue siendo el mismo. (Los verdaderos valores de w y z cambiarían pero la razón $w:z$ permanecería constante.) Cada uno de los vectores

$$|\psi\rangle,\ 2|\psi\rangle, -|\psi\rangle,\ i|\psi\rangle,\ \sqrt{2}\,|\psi\rangle, \pi\,|\psi\rangle,\ (1-3i)|\psi\rangle\ \text{, etcétera,}$$

representa el *mismo* estado físico, como lo hace cualquier $z|\psi\rangle$, con $z \neq 0$, y el único elemento del espacio de Hilbert que *no* tiene interpretación como un sistema físico es el vector *nulo* 0 (o el *origen* del espacio de Hilbert).

Para tener algún tipo de representación geométrica de todo esto, consideremos primero el concepto más usual de un vector "real". Normalmente visualizamos este vector simplemente como una *flecha* dibujada en un plano o en un espacio tridimensional. La suma de dos de estas flechas se obtiene mediante la ley del paralelogramo (véase fig. VI. 19). La operación de multiplicar un vector por un número (real) se obtiene, en términos de la representación de "flechas", multiplicando simplemente la longitud de la flecha por el número en cuestión y manteniendo constante la dirección de la misma. Si el número por el que multiplicamos es negativo, entonces se invierte su dirección, o si el número es cero, obtenemos el vector 0 que no tiene dirección. (El vector 0 se representa por la "flecha nula" de longitud cero.)

Un ejemplo de cantidad vectorial es la fuerza que actúa sobre una partícula. Otros ejemplos son las velocidades, aceleraciones y momentos clásicos. También están los cuadrivectores momento, que consideramos al final del último capítulo. Estos eran vectores en *cuatro* dimensiones en lugar de dos o tres. Sin embargo, para un espacio de Hilbert necesitamos vectores en dimensiones aún mucho mayores (a menudo infinitas, pero ésta no va a ser una consideración importante aquí). Recuérdese que también se utilizaban flechas para representar vectores en el espacio de fases clásico —que ciertamente podían ser de dimensión muy alta—. Las "dimensiones" en un espacio de fases no representan direcciones espaciales ordinarias, y tampoco lo hacen las "dimensiones" de un espacio de Hilbert. En lugar de ello, cada dimensión de un espacio de Hilbert corresponde a uno de los diferentes estados físicos independientes de un sistema cuántico.

Debido a la equivalencia entre $|\psi\rangle$ y $z|\psi\rangle$, un estado físico corresponde a una *línea a través del origen* 0, o *rayo,* en el espacio de Hilbert (descrita por todos los múltiplos de algún vector), y no simplemente a un vector

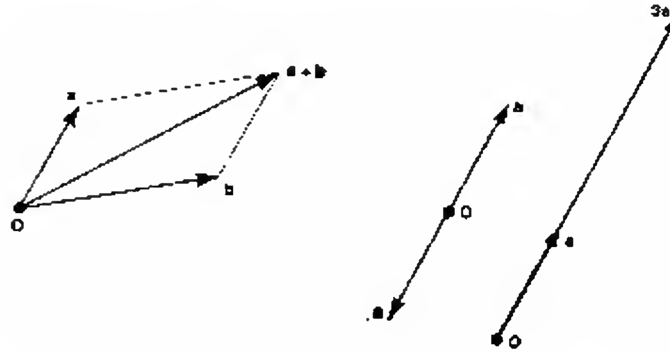


FIGURA VI. 19. *La suma de vectores en el espacio de Hilbert y la multiplicación por escalares pueden representarse normalmente, como en el caso de los vectores en el espacio ordinario.*

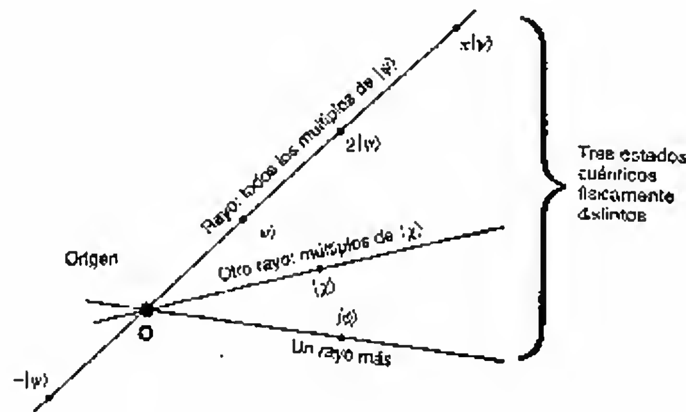


FIGURA VI.20. *Cada estado cuántico físico está representado por un rayo completo en el espacio de Hilbert.*

particular en dicha línea. El rayo consiste en todos los múltiplos posibles de un vector de estado particular $|\psi\rangle$. (Téngase en cuenta que son múltiplos *complejos,* de modo que la línea es en realidad una línea *compleja,* pero no nos preocupemos por esto ahora.) (Véase fig. VI.20).

Más adelante encontraremos una elegante representación de este espacio de rayos para el caso de un espacio de Hilbert *bi*dimensional. Un espacio de Hilbert de infinitas dimensiones aparece incluso en el caso sencillo de la localización de una sola partícula. Existe entonces una dimensión por cada una de las posibles posiciones que pudiera tener la partícula, y cada posición de la partícula define un "eje de coordenadas" en el espacio de Hilbert, de modo que con infinitas posiciones individuales diferentes para la partícula tenemos infinitas direcciones independientes (o dimensiones) diferentes en el espacio de Hilbert. Los estados de momento se representarán también en este *mismo* espacio de Hilbert y pueden presentarse como combinaciones de estados de posición, de manera que cada estado de momento corresponde a un eje en "diagonal", el cual está inclinado respecto a los ejes del espacio de posición. El conjunto de todos los estados de momento proporciona un nuevo conjunto de ejes, y el paso de los ejes del espacio de posición al de los ejes del espacio de momento implica una *rotación* en el espacio de Hilbert.

No necesitamos entender esto de manera precisa, pero algunas ideas tomadas de la geometría euclidiana ordinaria nos serán muy útiles. En particular, los ejes que hemos considerado *(o bien*

todos los ejes en el espacio de posición *o bien* todos los ejes en el espacio de momentos) deben considerarse como mutuamente *ortogonales,* es decir, que forman "ángulos rectos" entre sí.

La "ortogonalidad" entre rayos es un concepto importante para la mecánica cuántica: rayos ortogonales son estados *independientes* uno de otro. Los diferentes estados de posición posibles de una partícula son todos mutuamente ortogonales, como lo son todos los diferentes estados de momento posibles.

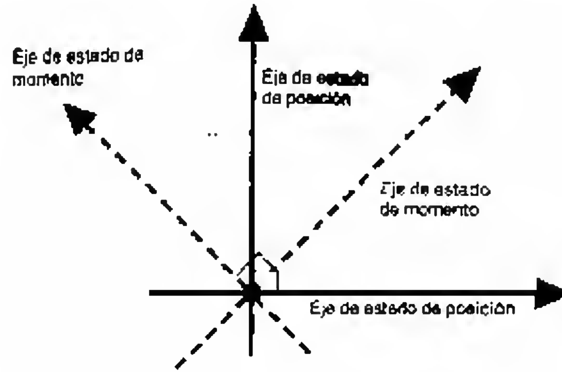


FIGURA VI.21. *Estados de posición y estados de momento proporcionan diferentes elecciones de ejes ortogonales en el mismo espacio de Hilbert.*

Pero los estados de posición no son ortogonales a los estados de momento. La situación se ilustra, muy esquemáticamente, en la fig. VI.21.

MEDIDAS

Como regla general R, en cada *medición* (u observación) los diferentes aspectos de un sistema cuántico que puedan amplificarse simultáneamente hasta el nivel clásico, y entre los cuales el sistema debe entonces escoger, deberán asimismo ser siempre *ortogonales.* Para una medición *completa,* las opciones seleccionadas constituirán un conjunto de vectores *base* ortogonales, lo que significa que cualquier vector del espacio de Hilbert puede ser (unívocamente) expresado linealmente en los mismos términos que los vectores base ortogonales. Para una medida *de posición* —en un sistema que conste de una sola partícula— esos vectores base definirán los ejes. Para el *momento* habrá un conjunto diferente que definirá los ejes del momento, y a cada tipo diferente de medida completa le corresponderá un conjunto propio.

Tras la medida, el estado del sistema *saltará* a uno de los ejes del conjunto determinado por la medida, siendo la nueva probabilidad la que gobierna la elección. No hay ley dinámica que nos diga cuál de entre los ejes seleccionados elegirá la naturaleza. Su elección es aleatoria y los valores de tal probabilidad serán obtenidos mediante los cuadrados de los módulos de las amplitudes respectivas.

Supongamos que se hace una medición completa sobre un sistema cuyo estado es $|\psi\rangle$, siendo los vectores base para la medida seleccionada

$$|1\rangle, |2\rangle, |3\rangle, |4\rangle, \ldots$$

Puesto que éstos forman un conjunto completo, cualquier vector de estado, y en particular $|\psi\rangle$, puede representarse linealmente* en los siguientes términos:

Geométricamente, las componentes z_0, z_1, z_2,... miden los *tamaños de las proyecciones ortogonales* del vector $|\psi\rangle$ sobre los diversos ejes $|0\rangle$, $|1\rangle$, $|2\rangle$, (Véase fig. VI.22).

Nos gustaría poder interpretar los números complejos z_0, z_1 z_2,...como las amplitudes de probabilidad que buscamos, de modo que los cuadrados de sus módulos proporcionen las diversas probabilidades de que el sistema se encuentre, tras la medición, en los respectivos estados $|0\rangle$, $|1\rangle$, $|2\rangle$,... Pero esto no basta porque no hemos fijado las "escalas" de los diversos vectores base $|0\rangle$, $|1\rangle$, $|2\rangle$,... Para ello, debemos especificar que son, en cierto sentido, *vectores unitarios* (esto es, vectores de "longitud" unidad) y, por lo tanto, forman los que en terminología matemática se llama una *base ortonormal* (de vectores mutuamente ortogonales y *normalizados* a la unidad).[6] Si $|\psi\rangle$ está también normalizado para ser un

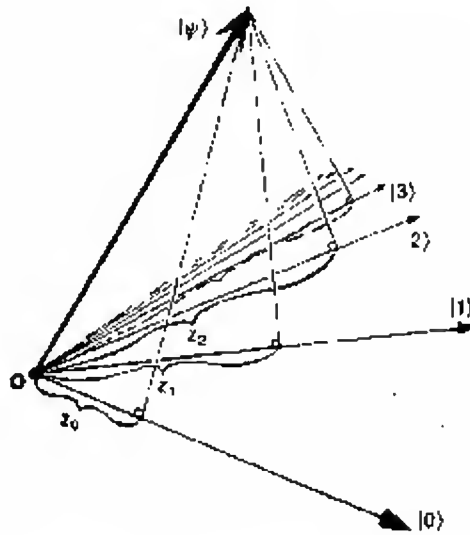


FIGURA VI.22. *Los tamaños de las proyecciones ortogonales del estado* $|\psi\rangle$ *sobre los ejes* $|0\rangle$, $|1\rangle$, $|2\rangle$,... *proporcionan las amplitudes requeridas* z_0, z_1, z_2,...

* Esto debe considerarse en el sentido de que se permite una suma *infinita* de vectores. La definición *completa* de un espacio de Hilbert es demasiado técnica para entrar ahora en ella, pero permite sumas infinitas de vectores.

6 Hay una operación importante, conocida como *producto escalar* (o producto interno) de dos vectores, que puede ser utilizada para expresar los conceptos de "vector unitario", "ortogonalidad" y "amplitud de probabilidad" de un modo muy simple. (En el álgebra vectorial ordinaria el producto escalar es *ab cos* θ , donde *a y b* son las longitudes de los vectores y θ es el ángulo entre sus direcciones.) El producto escalar entre vectores del espacio de Hilbert da un número *complejo*. Para dos vectores de estado $|\psi\rangle$ y $|\chi\rangle$ escribimos el producto escalar $\langle\psi|\chi\rangle$. Existen reglas algebraicas $\langle\psi|(|\chi\rangle + |\varphi\rangle)\rangle = \langle\psi|\chi\rangle + \langle\psi|\varphi\rangle$, $\langle\psi|(q|\chi\rangle)\rangle = q\langle\psi|\chi\rangle$,. y $\langle\psi|\chi\rangle = \overline{\langle\psi|\chi\rangle}$ donde la barra denota conjugación compleja. (El complejo conjugado de $z = x + iy$ es $\bar{z} = x - iy$, siendo x e y reales; nótese que $|z|^2 = z\bar{z}$) La ortogonalidad entre $|\psi\rangle$ y $|\chi\rangle$ se expresa como $\langle\psi|\chi\rangle = 0$ y . El cuadrado de la longitud de $|\psi\rangle$ es $|\psi|^2 = \langle\psi|\psi\rangle$, de modo que la condición para que $|\psi\rangle$ esté normalizada como un vector unitario es $\langle\psi|\psi\rangle = 1$ .Si un "acto de medida" provoca que un estado $|\psi\rangle$ salte o bien a $|\chi\rangle$ o a algún otro ortogonal a $|\chi\rangle$, entonces la amplitud de que salte a $|\chi\rangle$ es $\langle\chi|\psi\rangle$, suponiendo que $|\psi\rangle$ y $\chi\rangle$ estén ambos normalizados. Si no están normalizados, la probabilidad de saltar de $|\psi\rangle$ a $|\chi\rangle$ puede escribirse $\langle\chi|\psi\rangle\langle\psi|\chi\rangle / \langle\chi|\chi\rangle\langle\psi|\psi\rangle$ (Véase Dirac, 1947.)

vector unitario, entonces las amplitudes requeridas serán las componentes z_0, z_1, z_2,...de $|\psi\rangle$, y las respectivas probabilidades requeridas serán $|z_0|^2$, $|z_1|^2$, $|z_2|^2$..... Si $|\psi\rangle$ no es un vector unitario, entonces estos números *serán proporcionales* a las requeridas amplitudes y probabilidades en cada caso. Las verdaderas amplitudes serán

$$\frac{z_0}{|\psi|}, \frac{z_1}{|\psi|}, \frac{z_2}{|\psi|}, \text{etc}$$

y las verdaderas probabilidades

$$\frac{|z_0|^2}{|\psi|^2}, \frac{|z_1|^2}{|\psi|^2}, \frac{|z_2|^2}{|\psi|^2}, \text{etc}$$

donde $|\psi|$ es la longitud del vector de estado $|\psi\rangle$.

Esta "longitud" es un número real positivo definido para cada vector de estado (0 tiene longitud cero), y $|\psi| = 1$ si $|\psi\rangle$ es un es un vector unitario

Una medición completa es en realidad algo muy idealizado. La medición completa de la posición de una partícula, por ejemplo, exigiría que fuéramos capaces de localizarla, con precisión infinita, en cualquier parte del Universo. Un tipo de medida más elemental es uno en el que simplemente planteamos una pregunta *sí/no* tal como: "¿la partícula está a la izquierda o a la derecha de cierta línea?" O "¿está el momento de la partícula dentro de cierto intervalo?", etc. Las medidas *sí/no* son en realidad las más utilizadas. (Podemos, por ejemplo, estrechar tanto como queramos el margen de la posición de una partícula o del momento empleando sólo las medidas *sí/no*.) Supongamos que el resultado de una medición *sí/no* resulta ser "sí". Entonces el vector de estado debe encontrarse en la región "sí" del espacio de Hilbert, que llamaré **S**. Si, por el contrario, el resultado de la medida es "no", entonces el vector de estado se encuentra en la región "no" del espacio de Hilbert, que llamaré **N**. Las regiones **S** y **N** son completamente ortogonales entre sí, en el sentido de que cualquier vector de estado perteneciente a S debe ser ortogonal a cualquier vector de estado perteneciente a **N** (y *viceversa).* Además, cualquier otro vector de estado $|\psi\rangle$ puede ser expresado (de forma única) como una suma de vectores, uno de cada una de las **S** y **N**. En terminología matemática decimos que **S** y **N** son *complementos ortogonales* uno de otro. Así, $|\psi\rangle$ se expresa unívocamente como

$$|\psi\rangle = |\psi_S\rangle + |\psi_N\rangle$$

donde $|\psi_S\rangle$ pertenece a **S** y $|\psi_N\rangle$ pertenece a **N**. Aquí $|\psi_S\rangle$ es *la. proyección ortogonal* del estado $|\psi\rangle$ sobre **S** y, análogamente, $|\psi_N\rangle$ es la proyección ortogonal de $|\psi\rangle$ sobre **N**. (Véase fig. VI.23).

Durante la medición, el estado $|\psi\rangle$ *salta* y se convierte en (proporcional a) $|\psi_S\rangle$, o en $|\psi_N\rangle$. Si el resultado es "sí", entonces salta a $|\psi_S\rangle$ y si es "no", salta a $|\psi_S\rangle$. Si $|\psi\rangle$ está normalizado, las probabilidades respectivas de estas dos ocurrencias son las *longitudes al cuadrado*

$$|\psi_S|^2, |\psi_N|^2$$

de los estados proyectados.

Si $|\psi\rangle$ no está normalizada, dividiremos cada una de estas expresiones por $|\psi|^2$. (El teorema de Pitágoras, $|\psi|^2 = |\psi_S|^2 + |\psi_N|^2$ - asegura que la suma de estas probabilidades es la unidad.) Nótese

que en esta proyección la probabilidad de que $|\psi\rangle$ salte a $|\psi_S\rangle$ viene dada por la razón en que se reduce el cuadrado de su longitud.

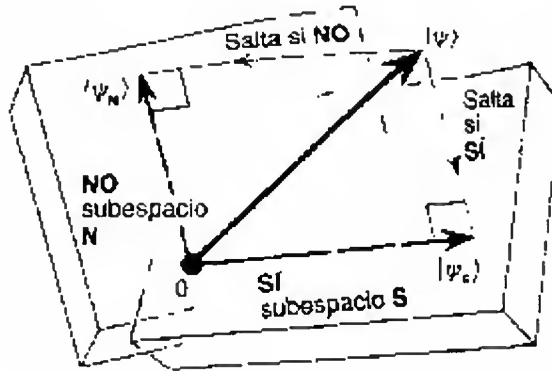


FIGURA VI.23. *Reducción del vector de estado. Una medición* sí/no *puede describirse en términos de un par de subespacios* **S** *y* **N** *que son complementos ortogonales entre sí. En la medición el estado* $|\psi\rangle$ *salta a su proyección sobre uno u otro de estos subespacios, con probabilidades dadas por el factor en que decrece el cuadrado de la longitud del vector de estado en la proyección.*

Debe hacerse una puntualización final concerniente a los "actos de medir" que pueden realizarse en un sistema cuántico. Una consecuencia de los principios de la teoría es que para *un estado cualquiera* —digamos el estado $|\chi\rangle$— puede realizarse en principio una medición *sí/no*[7] para la que en principio la respuesta es "sí" si el estado medido es (proporcional a) $|\chi\rangle$ y "no" si es ortogonal a $|\chi\rangle$. La región **S** anterior podría consistir, de este modo, en todos los múltiplos de un estado escogido $|\chi\rangle$. Esto implica que los vectores de estado son *reales objetivamente.* Cualquiera que sea el estado del sistema físico —y llamaremos $|\chi\rangle$ a dicho estado— existe una medición que puede ser realizada en principio para la que $|\chi\rangle$ es el *único* estado (salvo proporcionalidad) para el que la medición da el resultado "sí", con *certeza.* Para algunos estados $|\chi\rangle$ esta medición sería extremadamente difícil de realizar —quizá "imposible" en la práctica—, pero el hecho de que *en principio,* de acuerdo con la teoría, esta medida puede hacerse, trae consigo algunas consecuencias sorprendentes para nuestra discusión posterior.

EL SPIN Y LA ESFERA DE ESTADOS DE RIEMANN

El *spin* se considera a veces como "la más" cuántica de todas las cantidades físicas, y es prudente que le prestemos atención. ¿Qué *es* el spin? En esencia, es una medida de la rotación de una partícula. El término "spin" sugiere algo como el giro de una bola de cricket o de béisbol. Recordemos el concepto de *momento angular* que, como la energía y el momento, es una magnitud que se *conserva* (véase capítulo V). El momento angular de un cuerpo persiste mientras el cuerpo no sea perturbado por fuerzas de rozamiento o de otro tipo. En esto consiste, en realidad, el spin cuántico. Pero ahora lo que nos interesa es el giro de *una sola* partícula, no el

[7] Para aquellos familiarizados con el formalismo de los operadores de la mecánica cuántica, esta medida se define (en la notación de Dirac) por el operador hermítico acotado $|\chi\rangle\langle\chi|$. El valor propio 1 (para $|\chi\rangle$ normalizado) significa "sí" y el valor propio 0 significa "no". (Los vectores $\langle\chi|$, $\langle\psi|$, etc., pertenecen al espacio *dual* del espacio de Hilbert original.) Véase von Neumann (1955), Dirac (1947).

movimiento orbital de miríadas de partículas individuales en torno a su centro de masas común (que sería el caso de una bola de cricket). Un hecho físico notable es que la mayoría de las partículas encontradas en la naturaleza realmente gira en este sentido, con una cantidad específica para cada tipo de partícula.[8] No obstante, el spin de una simple partícula cuántica tiene propiedades peculiares que no son ni mucho menos las que esperaríamos a partir de nuestras experiencias con giros de bolas de cricket o cosas similares.

En primer lugar, la *magnitud* del spin de una partícula es siempre la *misma* para ese tipo concreto de partícula. Solamente la dirección del eje de giro es la que puede llegar a variar (de una muy extraña forma que ya abordaremos). Esto está en absoluto contraste con una bola de cricket, que puede girar en cualquier dirección y a diferentes velocidades, de acuerdo con la manera como sea lanzada. Para un electrón, protón o neutrón la cantidad de spin es siempre h/2, o sea únicamente la *mitad* del menor valor positivo que admitía originalmente Bohr para sus momentos angulares cuantizados de los átomos. (Recordemos que estos valores eran 0, h, 2h, 3h,...) Aquí requerimos la mitad de la unidad básica h, y, en cierto sentido, h/2 es ella misma la unidad básica fundamental. Esta cantidad de momento angular no estaría permitida para un objeto compuesto de un cierto número de partículas orbitando sin que ninguna de ellas estuviese girando sobre sí misma. Sólo aparece porque el spin es una propiedad *intrínseca* de la propia partícula (es decir, que no surge del movimiento orbital de sus "partes" en torno a su centro).

Una partícula cuyo spin es un múltiplo *impar* de h/2 (es decir, h/2, 3h/2, o 5h/2, etc.) se llama *fermión,* y exhibe una curiosa rareza de la descripción cuántica: una rotación completa de 360° transforma su vector de estado no en sí mismo sino en *menos* sí mismo. La mayoría de las partículas de la naturaleza son realmente fermiones, y más adelante oiremos más sobre ellas y sus singulares maneras de existir — tan vitales para nosotros — . Las partículas restantes, para las que el spin es un múltiplo *par* de h/2, es decir, un múltiplo entero de h (a saber 0, h, 2h, 3h,*...),* se llaman *bosones.* Tras una rotación de 360° el vector de estado de un bosón vuelve a sí *mismo,* no a su negativo.

Consideremos una partícula de spin 1/2, esto es con un valor h/2 para el spin. Para ser más concreto me referiré a la partícula como un *electrón,* pero un protón o un neutrón, o incluso un tipo adecuado de átomo, servirían igual. (Se admite que una "partícula" puede poseer partes individuales con tal de que pueda ser tratada cuánticamente como un todo simple, con un momento angular total bien definido.)

Tomemos el electrón en reposo y consideremos sólo su estado de spin. El espacio de estados cuánticos (espacio de Hilbert) resulta ser ahora *bi*dimensional, de modo que podemos tomar una base de sólo *dos* estados. Representaré estos estados como $|\uparrow\rangle$ y $|\downarrow\rangle$, con el propósito de indicar que para $|\uparrow\rangle$ el spin gira hacia la derecha, alrededor de la dirección vertical hacia *arriba,* mientras que para $|\downarrow\rangle$, gira hacia la derecha también, pero alrededor de la dirección hacia *abajo* (fig. VI.24). Los estados $|\uparrow\rangle$ y $|\downarrow\rangle$ son mutuamente ortogonales y, suponemos, normalizados ($|\uparrow|^2 = |\downarrow|^2 = 1$). Cualquier posible estado de spin del electrón es una superposición lineal,

---

[8] En mis primeras descripciones de un sistema cuántico consistente en una sola partícula he hecho demasiadas simplificaciones al ignorar el spin y suponer que el estado puede describirse en términos de su posición únicamente. *Existen* realmente ciertas partículas —llamadas partículas *escalares,* de las que son ejemplo las partículas nucleares llamadas *piones*, o ciertos átomos— para las que el valor del spin resulta ser cero. Para estas partículas (pero sólo para éstas) será suficiente la descripción anterior en términos de posición únicamente.

digamos $w|\uparrow\rangle + z|\downarrow\rangle$, de sólo los *dos* estados ortonormales $|\uparrow\rangle$ y $|\downarrow\rangle$, es decir de *arriba* y *abajo.*

Ahora bien, no hay nada especial en las direcciones "arriba" y "abajo". Exactamente igual podríamos haber descrito el spin girando en cualquier otra dirección (es decir, hacia la derecha o hacia la izquierda) cualquier otra dirección, pongamos por caso derecha $|\rightarrow\rangle$ como opuesto a izquierda $|\leftarrow\rangle$. Entonces (para una elección adecuada de la escala compleja para $|\uparrow\rangle$ y $|\downarrow\rangle$), encontramos *

$$|\rightarrow\rangle = |\uparrow\rangle + |\downarrow\rangle \text{ y } |\leftarrow\rangle = |\uparrow\rangle - |\downarrow\rangle$$

Lo que nos da una nueva visión: cualquier estado de spin electrónico es una superposición lineal de los dos estados ortogonales $|\rightarrow\rangle$ y $|\leftarrow\rangle$, es decir, de derecha e izquierda. Podríamos escoger, en su lugar, alguna dirección completamente arbitraria, por ejemplo la dada por el vector de



FIGURA VI.24. *Una base para los estados de spin de un electrón consta de sólo dos estados. Éstos pueden tomarse como* spin hacia arriba *y* spin hacia abajo.

estado |‰⟩. Esta es, nuevamente, una cierta combinación lineal compleja de $|\uparrow\rangle$ y $|\downarrow\rangle$. Digamos que

$$|‰\rangle = w|\uparrow\rangle + z|\downarrow\rangle$$

y todo estado de spin será una superposición lineal de ese estado y el estado ortogonal |Â⟩, que apunta en dirección opuesta[9] a |‰⟩. (Nótese que el concepto de "ortogonal" en el espacio de Hilbert no corresponde necesariamente a la formación de "ángulos rectos" en el espacio ordinario. En estos casos, los vectores ortogonales del espacio de Hilbert corresponden más a direcciones diametralmente opuestas en el espacio, que a direcciones que forman ángulos rectos.)

¿Cuál es la relación geométrica entre la dirección que determina |‰⟩ en el espacio y los dos números complejos *w* y *z*? Puesto que el estado físico dado por |‰⟩ queda inalterado si multiplicamos |‰⟩ por un número complejo distinto de cero, únicamente la razón de *z* a *w* será la realmente significativa.

Escribamos

$$q = z/w$$

* Como hice antes, prefiero no complicar las descripciones con factores como $1/\sqrt{2}$, que aparecerían si exigiésemos que $|\rightarrow\rangle$ y $|\leftarrow\rangle$ estén normalizados.

[9] Tómese $|Â\rangle = \bar{z}\,|\uparrow\rangle - \bar{w}\,|\downarrow\rangle$ donde $\bar{z}$ y $\bar{w}$ son los complejos conjugados de z y w (Véase nota 6.)

para este cociente. Aquí q corresponderá simplemente a algún número complejo, excepto porque el valor "$q = \infty$" también está permitido cuando hay que cubrir la situación $w = 0$, es decir, cuando la dirección del spin es la vertical hacia abajo. A menos que $q = \infty$, podemos representar q como un punto en el plano de Argand, igual que hicimos en el capítulo III.

Imaginemos que este plano de Argand está situado horizontalmente en el espacio, con el eje real en la dirección hacia la "derecha" en la descripción anterior (esto es, en la dirección del estado de spin $|\rightarrow\rangle$). Imaginemos una esfera de radio unidad cuyo centro está en el origen de este plano de Argand, de modo que los puntos 1, i, -1, -i están todos en el ecuador de la esfera. Consideremos el punto en el Polo Sur, que designaremos ∞. Proyectemos entonces desde este punto, de modo que el plano de Argand entero se aplique sobre la esfera. Así, cualquier punto q en el plano de Argand corresponderá a un punto q único en la esfera, obtenido al alinear los dos puntos con el Polo Sur (fig. VI.25). Esta correspondencia se llama proyección *estereográfica* y tiene muchas propiedades geométricas hermosas (por ejemplo, conserva los ángulos y aplica círculos en círculos).

La proyección nos da una caracterización de los puntos de la esfera con base en números complejos junto con ∞, mediante el conjunto de razones complejas q posibles. Una esfera caracterizada de esta forma particular se llama *esfera de Riemann.* Su importancia —para los estados de spin del electrón— reside en que la dirección del spin definido por $|\%\rangle = w|\uparrow\rangle + z|\downarrow\rangle$ viene dada por la dirección real desde el centro al punto $q = z/w$ marcado en la esfera de Riemann. Notemos que el Polo Norte corresponde al estado $|\uparrow\rangle$, que está dado por $z = 0$, es decir por $q = 0$, y el Polo Sur a $|\downarrow\rangle$, dado por $w = 0$, es decir por $q = \infty$. El punto más extremo hacia la derecha está caracterizado por $q = 1$, que proporciona el estado $|\rightarrow\rangle = |\uparrow\rangle + |\downarrow\rangle$, y el punto más extremo a la izquierda por $q = -1$, que proporciona el estado $|\leftarrow\rangle = |\uparrow\rangle - |\downarrow\rangle$

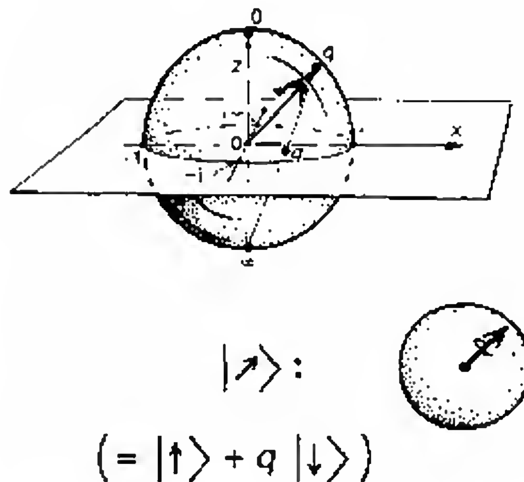


FIGURA VI.25. *La esfera de Riemann, representada aquí como el espacio de estados físicamente distintos de una partícula de spin 1/2. La esfera se proyecta estereográficamente desde su Polo Sur* (∞) *en el plano de Argand que pasa por su ecuador.*

El punto más alejado hacia el fondo de la esfera esta caracterizado por $q = i$, correspondiente al estado $|\uparrow\rangle + i|\downarrow\rangle$, en el que el spin apunta alejándose de nosotros, y el punto más próximo, $q = -i$, corresponde a $|\uparrow\rangle - i|\downarrow\rangle$, en el que el spin apunta directamente hacia nosotros. El punto más general, caracterizado por q, corresponde a $|\uparrow\rangle + q|\downarrow\rangle$

¿Cómo enlaza todo esto con las mediciones que pudiéramos realizar sobre el spin del electrón?[10] Seleccionemos alguna dirección en el espacio; y llamémosla dirección α . Si medimos el spin del electrón en esta dirección, la respuesta "sí" indica que el electrón está girando hacia la derecha, alrededor de α, mientras que "no" dice que gira en la dirección opuesta.

Ahora supongamos que la respuesta es "sí"; y caractericemos el estado resultante así: $|\alpha\rangle$. Si repetimos la medición utilizando exactamente la misma dirección a que antes, entonces encontraremos que la respuesta es de nuevo "sí" (con 100% de probabilidad). Pero si para la segunda medición cambiamos la dirección a una nueva dirección β, entonces vamos a encontrar que hay una probabilidad más pequeña para la respuesta "sí", y que el estado salta ahora a $|\beta\rangle$. Por lo que existiría alguna posibilidad de que la respuesta a la segunda medición fuera "no", y que el estado saltara a la dirección opuesta a β. ¿Cómo calcular esa probabilidad? La respuesta está contenida en la receta dada al final de la sección previa. La probabilidad de "sí" resulta ser, para la segunda medición,

$$1/2(1 + \cos\theta),$$

donde θ es el ángulo entre las direcciones[11] de α y β .

Y la probabilidad de "no" para la segunda medición es, en consecuencia,

$$1/2(1 - \cos\theta),$$

Como podemos ver, si la segunda medición se realiza en ángulo recto con la primera, la probabilidad de cualquiera de los resultados es del 50% ($\cos 90° = 0$): entonces el resultado de la segunda medición será completamente aleatorio. Si el ángulo entre las dos mediciones es agudo, entonces la respuesta "sí" es más probable que "no". Si es obtuso, entonces "no" es más probable que "sí". En el caso extremo en que β es opuesta a α, las probabilidades se hacen de 0% para "sí" y de 100% para "no". Por lo que es seguro que el resultado de la segunda medición será el contrario del correspondiente a la primera. (Véase Feynman *et al,* 1965, para más información sobre el spin.)

La esfera de Riemann juega un papel fundamental (pero no siempre reconocido) en *cualquier* sistema cuántico de dos estados, describiendo el conjunto de estados cuánticos posibles (salvo proporcionalidad). Para una partícula de spin 1/2, su papel geométrico es particularmente evidente porque los puntos de la esfera corresponden a las posibles direcciones espaciales para el eje de giro. En otras situaciones es difícil observar ese papel. Consideremos un fotón que acaba de atravesar un par de rendijas, o que ha sido reflejado por un espejo semirreflectante. El estado del fotón es alguna combinación lineal tal como $|\psi_m\rangle + |\psi_i\rangle$, $|\psi_m\rangle - |\psi_i\rangle$,o $|\psi_m\rangle + i|\psi_i\rangle$ de dos

[10] Hay un dispositivo experimental estándar, conocido como aparato de Stern-Gerlach, que puede ser utilizado para medir los *spines* de átomos apropiados. Los átomos se proyectan en un haz que atraviesa un campo magnético fuertemente inhomogéneo, y la dirección de la inhomogeneidad del campo proporciona la dirección de la medida del *spin.* El haz se desdobla en dos (para un átomo de *spin* 1/2, o en más de dos haces para un *spin* mayor), un haz que da los átomos para los que la respuesta a la medida del spin es "sí" y el otro para el que la respuesta es "no". Por desgracia, hay razones técnicas, irrelevantes para nuestros propósitos, por las que este aparato no puede ser utilizado para la medida del *spin* electrónico y debe utilizarse un procedimiento más indirecto. (Véase Mott y Massey, 1965.) Por ésta y otras razones prefiero no ser muy concreto sobre el modo en que realmente se está midiendo el *spin.*

[11] El lector decidido puede ocuparse en verificar la geometría dada en el texto. Resulta más fácil si orientamos nuestra esfera de Riemann de modo que la dirección α sea "arriba" y la dirección β esté en el plano que determinan "arriba" y "derecha", esto es, dado por $q = \tan(\theta/2)$ *en* la esfera de Riemann, y luego usamos la receta $\langle\chi|\psi\rangle\langle\psi|\chi\rangle/\langle\chi|\chi\rangle\langle\psi|\psi\rangle$ para 1a probabilidad de saltar de $|\psi\rangle$ a $|\chi\rangle$. Véase nota 6.

estados $|\psi_m\rangle$ y $|\psi_i\rangle$, los cuales describen dos localizaciones distintas. La esfera de Riemann describe todavía el conjunto de posibilidades físicamente distintas, pero ahora sólo de manera *abstracta.* El estado $|\psi_m\rangle$ está representado por el Polo Norte (cima) y $|\psi_i\rangle$ por el Polo Sur (fondo). Entonces, $|\psi_m\rangle + |\psi_i\rangle$, $|\psi_m\rangle - |\psi_i\rangle$, y $|\psi_m\rangle + i|\psi_i\rangle$ están representados por los diversos puntos en el ecuador, y en general $w\,|\psi_m\rangle + z\,|\psi_i\rangle$ está representado por el punto dado por $q = z/w$. *En* muchos casos, como éste, el "valor de posibilidades de la esfera de Riemann" está bastante oculto, como oculta queda su relación con la geometría espacial.

OBJETIVIDAD Y MESURABILIDAD DE LOS ESTADOS CUÁNTICOS

Pese al hecho de que normalmente sólo disponemos de probabilidades para el resultado de un experimento, hay algo *objetivo* en un estado mecánico-cuántico. Se afirma con frecuencia que el vector de estado es simplemente una descripción convencional de "nuestro conocimiento" respecto de un sistema físico —o, tal vez, que el vector de estado no describe en realidad un sistema simple sino que únicamente proporciona información probabilística sobre un "conjunto" de un gran número de sistemas preparados de forma similar. Son opiniones injustificablemente tímidas si las comparamos con lo que la mecánica cuántica todavía tiene que decirnos sobre la *realidad* del mundo físico. Parte de esta duda respecto a la "realidad física" de los vectores de estado parece surgir del hecho de que lo que es físicamente mesurable está limitado de forma estricta, según la teoría.

Consideremos el estado de spin de un electrón, como los descritos antes. Supongamos que el estado de spin sea $|\alpha\rangle$, pero que nosotros no lo sabemos; es decir, nosotros no conocemos la *dirección* α del eje alrededor del que se supone está girando el electrón. ¿Podemos determinar esta dirección mediante una medición? No, no podemos. Lo más que podemos hacer es extraer "un poco" de información, es decir, la respuesta a una simple pregunta *sí/no.* Podemos seleccionar alguna dirección β en el espacio y medir el spin del electrón en dicha dirección. Obtendremos la respuesta "sí" o "no". Pero, desde ese instante habremos perdido la información sobre la dirección original del spin. Con una respuesta "sí" sabemos que el estado es *ahora* proporcional a $|\beta\rangle$, y con una respuesta "no" sabemos que el estado está *ahora* en la dirección opuesta a β. En ninguno de los dos casos eso nos dirá cuál es la dirección α del estado *antes* de la medición, por lo que habrá que contentarse con una mera información probabilística sobre α.

Por otra parte, parecería haber algo completamente *objetivo* sobre la propia dirección α en la que el electrón "estaría girando" antes de que se hiciese la medición.* En efecto, *podríamos* haber medido el spin del electrón en la dirección α, y el electrón tendría que haber estado preparado para dar la respuesta "sí" con *certeza,* si nuestra conjetura hubiera estado en el camino correcto. De algún modo, la "información" de que el electrón va a dar realmente tal respuesta está almacenada en el estado de spin del electrón.

Como sea, al discutir la cuestión de la realidad física, es necesario distinguir entre lo "objetivo" y lo "mesurable", según la mecánica cuántica. En efecto, el vector de estado de un sistema es *no mesurable* en el sentido de que no podemos verificar exactamente (salvo proporcionalidad), mediante experimentos realizados sobre el sistema, cuál es ese estado, y no obstante, el vector de

* Esta objetividad es una característica y resulta de nuestra aceptación del formalismo estándar de la mecánica cuántica. Desde un punto de vista *no* estándar, el sistema podría "saber" realmente, antes de tiempo, el resultado que daría en *cualquier* medición. Lo que podría darnos una imagen *diferente,* aparentemente objetiva, de la realidad física.

estado *parece* ser (de nuevo salvo proporcionalidad) una propiedad totalmente *objetiva* del sistema, toda vez que éste —el sistema— se caracteriza por los resultados que debe dar en los experimentos que *pudieran* realizarse. En el caso de una simple partícula de spin 1/2, como un electrón, esta objetividad es razonable porque lo que hace es simplemente afirmar que hay *alguna* dirección en la que el spin del electrón está exactamente definido, incluso aunque no podamos saber cuál es esa dirección. (Sin embargo, veremos más adelante que esta imagen "objetiva" se torna muy extraña con sistemas más complicados, incluso para un sistema que conste sólo de un *par* de partículas de spin 1/2.) Pero, ¿es necesario que el spin del electrón tenga un estado físicamente definido antes de ser medido? En muchos casos *no* lo tendrá, puesto que no puede considerarse como un sistema cuántico por sí mismo. En lugar de ello, el estado cuántico debe tomarse generalmente como la descripción de un electrón inextricablemente mezclado con gran número de partículas. En circunstancias particulares, no obstante, el electrón (al menos en lo que respecta a su spin) *puede* ser considerado independientemente. Por ejemplo, cuando un spin ha sido medido previamente en alguna dirección (quizá desconocida) y luego ha permanecido sin perturbación durante un tiempo, entonces el electrón *tiene* una dirección de spin perfecta u objetivamente definida, según la teoría cuántica estándar.

COPIA DE UN ESTADO CUÁNTICO

La objetividad pero no-mensurabilidad de un estado de spin del electrón ilustra otro hecho importante: *es imposible copiar un estado cuántico y dejar intacto el estado original.* Supongamos que pudiéramos hacer tal copia de un estado de spin del electrón $|\alpha\rangle$. Si pudiéramos hacerlo una vez, podríamos hacerlo otra vez, y otra y otra. El sistema resultante podría tener un enorme momento angular con una dirección muy definida. Tal dirección, a saber a, podría ser determinada mediante una medición macroscópica, lo cual violaría la *no* mensurabilidad fundamental del estado de spin $|\alpha\rangle$.

En cambio, *es* posible copiar un estado cuántico si al mismo tiempo estamos dispuestos a destruir el estado del original. Por ejemplo, podríamos tener un electrón en un estado de spin $|\alpha\rangle$ desconocido y un neutrón en otro estado de spin $|\gamma\rangle$. Es totalmente legítimo intercambiarlos, de modo que el estado de spin del neutrón es ahora $|\alpha\rangle$ y el del electrón es $|\gamma\rangle$. Lo que no podemos hacer es *duplicar* $|\alpha\rangle$ (a menos que ya *supiéramos* cuál es $|\alpha\rangle$ realmente). *(Cfr.* también Wootters y Zurek, 1982.)

Recordemos la "máquina teleportadora" discutida en el capítulo I. Esta depende de que sea posible, en principio, reconstruir una copia completa del cuerpo y el cerebro de una persona en un planeta distante. Resulta intrigante especular que la "conciencia" de una persona pueda depender de un estado cuántico. Si es así, la teoría cuántica nos prohibiría hacer una copia de esta "conciencia" sin destruir el estado del original y, de esta forma, la "paradoja" de la teleportación podría resolverse. La relación entre los efectos cuánticos y la función cerebral se considerará en los dos últimos capítulos.

EL SPIN DEL FOTÓN

Consideraremos a continuación el "spin" de un fotón y su relación con la esfera de Riemann. Los fotones *poseen* spin pero, debido a que siempre viajan a la velocidad de la luz, no podemos

considerar el spin como si estuviera en un punto fijo. En lugar de ello, el spin del fotón está siempre en la dirección del movimiento.

Al spin del fotón se le llama *polarización,* que es el fenómeno del que depende el comportamiento de los anteojos "polarizados" para sol. Tomen dos piezas de cristal polarizado puestas una sobre otra, miren al través de ellas y verán que dejan pasar cierta cantidad de luz. Ahora giren una de las dos piezas manteniendo la otra fija. La cantidad de luz que dejan pasar variará. En una cierta orientación, la superposición de la segunda pieza no resta prácticamente nada de la luz entrante. Mientras que orientándola de modo que forme un ángulo recto con la anterior, se reducirá la luz prácticamente a cero.

Lo que sucede se puede entender mejor en términos de la imagen ondulatoria de la luz. En este caso, necesitamos la descripción de Maxwell de los campos eléctricos y magnéticos oscilantes. En la fig. VI.26, se ilustra la *luz plano-polarizada.* El campo eléctrico oscila hacia adelante y hacia atrás en un plano —llamado *plano de polarización*— y el campo magnético oscila al unísono pero en un plano que forma un ángulo recto con el del campo eléctrico. Cada pieza polarizada deja pasar la luz cuyo plano de polarización está alineado con su propia estructura. Cuando las dos piezas tienen su estructura orientada de la misma forma, toda la luz que atraviese la primera, atravesará también la segunda, pero cuando las dos tienen sus estructuras formando ángulos rectos, la segunda bloquea toda la luz que dejaba pasar la primera. Si las dos están orientadas formando un ángulo φ entre ellas, entonces la segunda deja pasar una fracción

$$\cos^2 \varphi$$

En la imagen de partículas debemos pensar que *cada fotón individual* posee polarización. El primer cristal polarizado actúa como un medidor de polarización que da la respuesta "sí", si el fotón está polarizado en la dirección apropiada, y el fotón puede pasar. Si el fotón está polarizado en la dirección ortogonal, entonces la respuesta es "no" y el fotón es absorbido. (Aquí, "ortogonal" en el sentido del espacio de Hilbert *corresponde* a un "ángulo recto" en el espacio ordinario.)



FIGURA VI. 26. *Una onda electromagnética plano-polarizada.*

Cuando el fotón atraviesa el primer cristal polarizado, el segundo plantea la pregunta correspondiente pero para alguna otra dirección. Y puesto que el ángulo entre estas dos direcciones es φ, como antes, tenemos ahora una *probabilidad* $\cos^2 \varphi$ de que el fotón haya atravesado el segundo cristal una vez que atravesó el primero.

¿Dónde interviene la esfera de Riemann? Para obtener la colección completa de estados de polarización en forma compleja debemos considerar polarizaciones *circular* y *elíptica.* Para una onda clásica, éstas quedan ilustradas en la fig. VI.27. Con polarización circular el campo eléctrico *rota,* en lugar de oscilar, y el campo magnético rota al unísono formando siempre un ángulo recto con el campo eléctrico. Para la polarización elíptica existe una combinación de movimientos (oscilatorio y rotacional), y el vector que representa el campo eléctrico describe una

elipse en el espacio. En la descripción cuántica, cada fotón *individual* presenta estas diferentes formas de estar polarizado: los estados de *spin del fotón.*

Para ver cómo la colección de posibilidades es otra vez la esfera de Riemann, imaginemos que el fotón está viajando verticalmente hacia arriba. El Polo Norte representa ahora el estado $|D\rangle$ de spin hacia la *derecha,* lo que significa que a medida que pasa el fotón el vector de campo eléctrico rota alrededor de la vertical en sentido contrario a las agujas del reloj (tal como se vería desde arriba). El Polo Sur representa el estado $|I\rangle$ de spin hacia la *izquierda.* (Podemos imaginar el fotón girando como las balas de un rifle, hacia la derecha o hacia la izquierda.) El estado de spin general $|D\rangle + q\,|I\rangle$ es una combinación lineal compleja de los dos y corresponde a un punto, caracterizado por *q,* en la esfera de Riemann. Para encontrar la relación entre *q* y la. elipse de polarización, tomemos la *raíz cuadrada* de *q* para obtener otro número complejo *p:*

$$p=\sqrt{q}$$

En seguida marquemos *p* en lugar de *q* en la esfera de Riemann y consideremos el plano que pasa por el centro de la esfera y que es perpendicular a la línea recta que une el centro con el punto *q*. Este plano intersecta

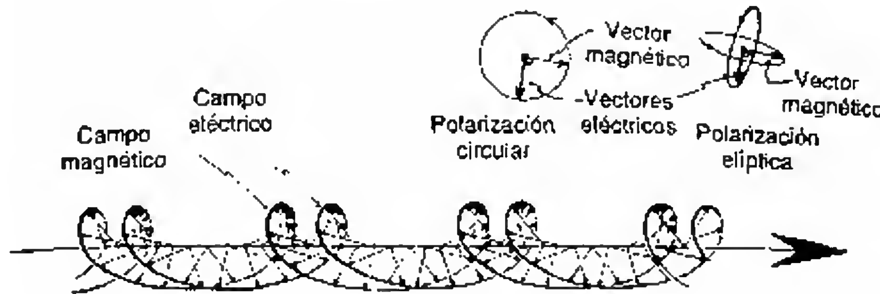


FIGURA VI.27. *Una onda electromagnética circularmente polarizada. (La polarización elíptica es intermedia, entre las figs. VI.26 y VI.27.)*

a la esfera en un círculo, así que proyectamos este círculo verticalmente hacia abajo para obtener la elipse de polarización (fig. VI.28).* La esfera de Riemann de los *q* describe aún la totalidad de los estados de polarización del fotón, pero la raíz cuadrada p de *q* proporciona su realización espacial.

Para calcular probabilidades podemos utilizar la misma fórmula 1/2(l + cosθ) que usamos para el electrón, a condición de que la apliquemos a *q* y no a *p.* Consideremos la polarización del plano. Midamos primero la polarización del fotón en una dirección y luego en otra, a un ángulo φ con la primera, direcciones que corresponden a dos valores de *p* en el ecuador de la esfera que subtienden φ en el centro. Puesto que las *p* son las raíces cuadradas de las *q,* el ángulo 0 subtendido en el centro por los puntos *q* es *doble* del que subtienden los puntos *p:* θ = 2φ . Así, la probabilidad de "sí" para la segunda medición, una vez que se obtuvo "sí" para la primera (es

* El número complejo *p* sería tan válido como *p* para la raíz cuadrada de *q, y* da la misma elipse de polarización. La raíz cuadrada tiene que ver con el hecho de que el fotón es una partícula sin masa y de *spin uno,* es decir, el *doble* de la unidad fundamental h/2. Para un *gravitón* —el aún no detectado cuanto de gravitación— el spin será *dos,* es decir, *cuatro* veces la unidad fundamental, y tendríamos que tomar la raíz *cuarta* de *q* en la descripción anterior.

decir, que el fotón atraviese el segundo cristal una vez que ha atravesado el primero) es 1/2(l+cosφ), que (por simple trigonometría) es lo mismo que el $\cos^2\varphi$ que resultaba antes.

OBJETOS CON GRAN SPIN

Para un sistema cuántico en que el número de estados base es mayor que dos, el espacio de estados físicamente distinguibles es más complicado que la esfera de Riemann.

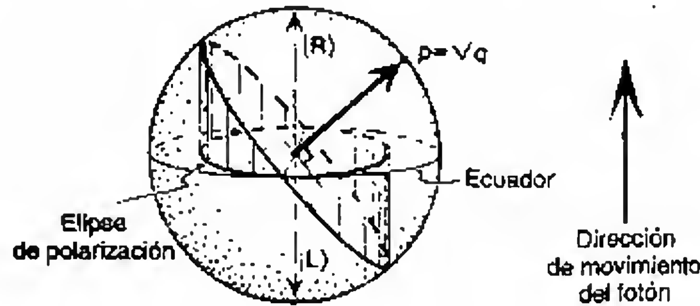


FIGURA VI.28. *La esfera de Riemann (pero ahora de* $\sqrt{q}$ *) describe también estados de polarización de un fotón. (El vector que apunta hacia* $\sqrt{q}$ *se llama vector de Stokes.)*

Sin embargo, en el caso del spin la propia esfera de Riemann tiene siempre por sí misma un papel geométrico directo que jugar. Consideremos una partícula o átomo en reposo *con masa* y de spin *n* x ħ/2. El spin define entonces un sistema cuántico de *(n* + 1) estados. (Para una partícula *sin masa* que gira, es decir, una que viaje a la velocidad de la luz, tal como un fotón, el spin es siempre un sistema de *dos* estados como el descrito antes, pero para una partícula con masa el número de estados crece con el spin.) Si decidimos medir este spin en alguna dirección, encontramos que existen *n* + 1 diferentes resultados posibles, dependiendo de la cantidad del spin que se halle orientada a lo largo de dicha dirección. En términos de la unidad fundamental ħ/2, los posibles resultados para el valor del spin en dicha dirección son *n, n -2, n* - 4,..., 2 - *n* o -*n.* Por lo cual, para *n* = 2 los valores son 2, 0 o -2; para *n* = 3, los valores son 3, 1, -1, o -3, etc. Los valores *negativos* corresponden al spin que apunta principalmente en la dirección *opuesta* a la que se está midiendo. El caso de spin 1/2, es decir n = 1, el valor 1 corresponde a "sí" *y* el valor -1 corresponde a "no".

Ahora bien, resulta, aunque no intentaré explicar las razones (Majorana, 1932; Penrose, *1987a),* que *todo estado de spin* (salvo proporcionalidad) para spin ħ*n*/2 está caracterizado unívocamente por un *conjunto* (no ordenado) *de* n *puntos en la esfera de Riemann* —es decir, por *n* direcciones (habitualmente distintas) que salen del centro hacia afuera (véase fig. VI.29). (Estas direcciones están caracterizadas por las mediciones que podríamos realizar en el sistema: si medimos el spin en una de ellas, es seguro que el resultado no estará completamente en la dirección opuesta, esto es, dará uno de los valores *n, n* - 2, *n* - 4,..., 2 - *n,* pero no -*n*.) En el caso particular *n* = 1, como el electrón anterior, tenemos *un* punto en la esfera de Riemann, *y* éste es simplemente el punto caracterizado por *q* en las descripciones que hicimos.

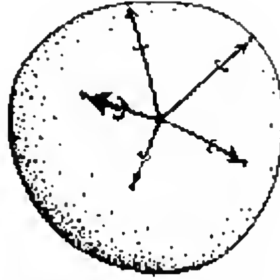


FIGURA VI.29. *Un estado general de spin superior para una partícula con masa, puede describirse como una colección de estados de spin 1/2 apuntando en direcciones arbitrarias.*

Pero para valores mayores del spin la imagen es más elaborada y es como la acabo de describir, aunque no es muy familiar a los físicos.

Hay algo enigmático en esta descripción. Frecuentemente se tiende a pensar que, en un sentido apropiado de límite, las descripciones cuánticas de los átomos (o las partículas elementales o las moléculas) coincidirán aproximadamente con las newtonianas clásicas cuando el sistema sea grande y complicado. Sin embargo, dicho así, esto es *sencillamente falso* porque —como hemos visto— los estados de spin de un objeto de gran momento angular corresponderán a un gran número de puntos salpicados por toda la esfera de Riemann.* Podemos considerar el spin del objeto como un lote completo de spines 1/2 apuntando en todas las direcciones que determinan estos puntos. Sólo algunos de esos estados combinados —a saber, aquellos en los que la mayoría de los puntos se concentran en una pequeña región de la esfera (es decir, en los que la mayoría de los spines y apuntan aproximadamente en la misma dirección)— corresponderán a los verdaderos estados de momento angular que encontramos normalmente con los objetos clásicos, como las bolas de cricket. Hubiéramos esperado que si escogemos un estado de spin para el que la medida total de éste es un número muy grande (en términos de $\hbar/2$), aunque *al azar* en todo lo demás, entonces empezara a emerger algo similar al spin clásico. Pero no es así como funcionan las cosas.

En general, los estados de spin cuánticos cuando el spin total es grande no se parecen en nada a los clásicos.

¿Cómo debe hacerse, entonces, la correspondencia con el momento angular de la física clásica?

Aunque la mayoría de los estados cuánticos de spin grande *no* se parecen a los clásicos, ellos son combinaciones lineales de estados (ortogonales) *cada uno* de los cuales se *parece* a uno clásico. De algún modo se realiza automáticamente una "medición" en el sistema y el estado "salta" (con cierta probabilidad) a uno u otro de esos estados similares a los clásicos. La situación es análoga con otras propiedades del sistema clásicamente medibles, y no sólo el momento angular. Es este aspecto de la mecánica cuántica el que debe entrar en juego cuando se quiere que un sistema "alcance el nivel clásico". Antes de que podamos discutir los sistemas cuánticos "grandes" o "complicados", tendremos que poseer alguna noción acerca de la manera singular en que la mecánica cuántica trata los sistemas que incluyen más de una partícula.

* Más exactamente, el momento angular se describe mediante una combinación lineal compleja de tales colecciones con diferentes números de puntos, y la razón es que en el caso de un sistema complicado puede haber superpuestos varios valores diferentes del spin total. Esto sólo hace que la imagen total sea aún *menos* parecida a la de un momento angular clásico.

SISTEMAS DE MUCHAS PARTÍCULAS

Las descripciones cuánticas de estados de muchas partículas son, complicadas, y pueden llegar a serlo en extremo. Debemos pensar en términos de superposiciones de *todas* las diferentes localizaciones posibles de todas las partículas por separado. Esto da un vasto espacio de estados posibles; mucho más que en el caso de un *campo* en la teoría clásica.

Hemos visto que incluso el estado cuántico de una partícula *simple,* a saber, una función de onda, tiene el mismo tipo de complicación que el de un campo clásico completo. Esta imagen (que requiere un número *infinito* de parámetros para especificarlo) es ya mucho más complicada que la imagen clásica de una partícula (que sólo necesita unos cuantos números para especificar su estado; en realidad seis, si no tiene grados de libertad internos, tales como spin; véase capítulo V). Esto puede parecer ya suficientemente grave, y podría pensarse que para describir un estado cuántico para dos partículas serían necesarios *dos* "campos": uno para describir el estado de cada una de las partículas. Nada de eso. Con dos o más partículas la descripción del estado es mucho más elaborada.

El estado cuántico de una partícula *simple* (sin spin) está definido por un número (amplitud) complejo para cada posible posición de la partícula. Ésta tiene una amplitud de estar en un punto A y una amplitud de estar en un punto B y una amplitud de estar en un punto C, etcétera.

Pensemos en *dos* partículas. La primera podría estar en A y la segunda en B. Tendría que haber una amplitud para esa posibilidad. A la inversa, la primera partícula podría estar en B y la segunda en A, y esto también necesitaría una amplitud. O la primera podría estar en B y la segunda en C o quizá ambas partículas podrían estar en A... Cada una de tales opciones necesitaría una amplitud propia. Así, la función de onda no es sólo un par de funciones de posición (es decir, un par de campos). Es una función de dos posiciones.

Para tener una idea de lo mucho más complicado que resulta especificar una función de dos posiciones, en comparación con lo que sería especificar dos funciones de una posición, imaginemos una situación en la que sólo hay disponible un conjunto finito de posiciones posibles. Supongamos que sólo hay diez posiciones permitidas, dadas por los estados ortonormales

$$|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle, |7\rangle, |8\rangle, |9\rangle.$$

El estado $|\psi\rangle$ de una simple partícula será alguna combinación .

$$|\psi\rangle = z_0|0\rangle + z_1|1\rangle + z_2|2\rangle + z_3|3\rangle + ... + z_9|9\rangle$$

en la que las diversas componentes $z_0, z_1, z_2, z_3, ..., z_9$ proporcionan las amplitudes correspondientes a cada uno de los puntos en que se encuentre la partícula. Diez números complejos especifican el estado de la partícula. Para un estado de *dos* partículas, necesitaremos una amplitud por cada *par de posiciones.* Existen $10^2 = 100$ diferentes pares (ordenados) de posiciones y por eso necesitamos *cien* números complejos. Si simplemente tuviéramos dos estados de una partícula (es decir, "dos funciones de posición" en lugar de "una función de dos posiciones") hubiéramos necesitado sólo *veinte* números complejos. Podemos caracterizar estos cien números complejos como

$$z_{00}, z_{01}, z_{02}, ..., z_{09}, z_{10}, z_{11}, z_{12}, ..., z_{20}, ..., z_{99}$$

y los correspondientes vectores base (ortonormales) como[12]

$$|0\rangle|0\rangle,\ |0\rangle|1\rangle,\ |0\rangle|2\rangle,\ldots\ldots,|0\rangle|9\rangle,\ |1\rangle|0\rangle,\ldots\ldots,|9\rangle|9\rangle.$$

Entonces, el estado general de dos partículas $|\psi\rangle$ tendría la forma

$$|\psi\rangle = z_{00}|0\rangle|0\rangle + z_{00}|0\rangle|1\rangle + \ldots + z_{00}|9\rangle|9\rangle$$

Esta notación de *producto* de estados tiene el significado siguiente: si $|\alpha\rangle$ es un posible estado para la primera partícula (no necesariamente un estado de posición), y si $|\beta\rangle$ es un estado posible para la segunda partícula, entonces el estado que afirma que el estado de la primera partícula es | $|\alpha\rangle$ *y* el de la segunda es $|\beta\rangle$ se escribirá así:

$$|\alpha\rangle|\beta\rangle$$

También pueden tomarse "productos" entre cualquier otro par de estados cuánticos, no necesariamente estados de partículas individuales. De ese modo, interpretamos siempre el estado producto $|\alpha\rangle|\beta\rangle$ (no necesariamente estados de partículas individuales) como cuando describimos la conjunción:

"el primer sistema tiene estado $|\alpha\rangle$ *y* el segundo sistema tiene estado $|\beta\rangle$

Una interpretación similar sería válida para $|\alpha\rangle|\beta\rangle|\gamma\rangle$, etc. Sin embargo, el estado *general* de dos partículas no tiene en realidad esta forma de "producto". Por ejemplo, podría ser

$$|\alpha\rangle|\beta\rangle+|\rho\rangle|\sigma\rangle$$

donde $|\rho\rangle$ es otro estado posible para el primer sistema y $|\sigma\rangle$, otro estado posible para el segundo. Este estado es una *superposición lineal.* A saber, la primera conjunción ($|\alpha\rangle$y$|\beta\rangle$) *más* la segunda conjunción ($|\rho\rangle$y$|\sigma\rangle$) no puede reexpresarse como un simple producto (es decir, como una conjunción de dos estados). Como un ejemplo más, el estado $|\alpha\rangle|\beta\rangle-i|\rho\rangle|\sigma\rangle$ describiría una superposición lineal diferente. Nótese que la mecánica cuántica exige mantener una clara distinción entre los significados de las palabras "y" y "más". Existe una desafortunada tendencia en el habla moderna — como en los folletos de seguros — a usar erróneamente "más" en el sentido de "y". Aquí debemos ser más cuidadosos.

La situación con tres partículas es similar. Para especificar un estado general de tres partículas, en el caso anterior en el que sólo había disponibles diez opciones para la posición, necesitamos ahora *mil* números complejos. La base completa para los estados de tres partículas será

$$|0\rangle|0\rangle|0\rangle,\ |0\rangle|0\rangle|1\rangle,\ |0\rangle|0\rangle|2\rangle,\ldots\ldots,|9\rangle|9\rangle|9\rangle,$$

Los estados específicos para las tres partículas tienen la forma

$$|\alpha\rangle|\beta\rangle|\gamma\rangle$$

(donde $|\alpha\rangle$, $|\beta\rangle$ y $|\gamma\rangle$ no necesitan ser estados de posición), pero para el estado general de tres partículas tendríamos que superponer muchos estados de este tipo simple de "producto". La pauta correspondiente para cuatro o más partículas deberá estar ahora clara.

[12] En lenguaje matemático decimos que el espacio de los estados de dos partículas es el *producto tensorial* del espacio de estados de la primera partícula por el de la segunda partícula. El estado $|\chi\rangle|\varphi\rangle$ es entonces el producto tensorial del estado $|\chi\rangle$ y el $|\varphi\rangle$.

Hasta ahora hemos examinado los casos de partículas *distinguibles,* en los que consideramos que la "primera partícula", la "segunda", la "tercera," etc., son todas de *diferentes* tipos. Sin embargo, la mecánica cuántica tiene la característica sorprendente de que para partículas *idénticas* las reglas son diferentes. De hecho, las reglas son tales que, en un sentido preciso, las partículas de un tipo específico tienen que ser *exactamente* idénticas y ya no sólo, por ejemplo, muy aproximadamente idénticas. Esto se aplica a todos los electrones y a todos los fotones. Pero, como debe ser, todos los electrones son idénticos entre sí de un modo *diferente* de como son idénticos todos los fotones. La diferencia estriba en el hecho de que los electrones son fermiones mientras que los fotones son bosones. Estos dos tipos generales de partículas tienen que ser tratados de forma diferente.

Antes de que confunda al lector, explicaré cómo deben ser caracterizados los estados fermiónicos como los estados bosónicos. La regla es como sigue: si $|\psi\rangle$ es un estado que incluye cierto número de fermiones de un tipo particular, al intercambiar dos de esos fermiones, $|\psi\rangle$ se transforma:

$$|\psi\rangle \rightarrow -|\psi\rangle$$

Si $|\psi\rangle$ incluye un número de bosones de un tipo particular, al intercambiar dos cualesquiera de esos bosones también se transforma $|\psi\rangle$, pero así:

$$|\psi\rangle \rightarrow |\psi\rangle$$

Una consecuencia de esto es que *dos fermiones no pueden estar en el mismo estado,* porque si lo estuvieran su intercambio no afectaría en absoluto al sistema total, de modo que deberíamos tener $-|\psi\rangle = |\psi\rangle$, es decir $|\psi\rangle = 0$, lo que no es posible en un estado cuántico. Esta propiedad se conoce como *principio de exclusión de Pauli,*[13] y tiene consecuencias fundamentales para la estructura de la materia. Los constituyentes principales de la materia son, efectivamente, fermiones: electrones, protones y neutrones. Sin el principio de exclusión la materia se concentraría y se colapsaría.

Examinemos de nuevo nuestras diez posiciones y supongamos que tenemos un estado que consiste en dos fermiones idénticos. El estado $|0\rangle|0\rangle$ está excluido por el principio de Pauli (si intercambiamos el primer factor con el segundo, el estado vuelve a sí mismo en lugar de a su negativo). Además, $|0\rangle|1\rangle$ no puede existir en esta forma porque bajo el intercambio no se transforma en su negativo, pero esto se remedia reemplazándolo por

$$|0\rangle|1\rangle - |1\rangle|0\rangle$$

(Podría incluirse también un factor global $1/\sqrt{2}$ con fines *de* normalización.)

Este estado cambia correctamente de signo bajo un intercambio de la primera partícula con la segunda, pero ahora no tenemos $|0\rangle\ |1\rangle$ y $|1\rangle\ |0\rangle$ como estados independientes. En lugar de aquellos *dos* estados, ahora sólo se permite *un* estado.

[13] Wolfgang Pauli, un brillante físico austríaco y figura prominente en el desarrollo de la mecánica cuántica, propuso su principio de exclusión como una hipótesis en 1925. El tratamiento mecánico-cuántico de lo que ahora llamamos "fermiones" fue desarrollado en 1926 por el muy influyente y original físico italoestadounidense Enrico Fermi y por el gran Paul Dirac, a quien ya hemos encontrado en varias ocasiones antes de ahora. El comportamiento estadístico de los fermiones sigue la "estadística de Fermi-Dirac", cuyo nombre la distingue de la "estadística de Boltzmann — la estadística clásica de las partículas distinguibles—. La "estadística de Bose-Einstein" para los bosones fue desarrollada para el tratamiento de los fotones por el famoso físico indio S. N. Bose y Albert Einstein en 1924.

En total hay

$$1/2(10x9) = 45$$

estados de este tipo, uno por cada par no ordenado de diferentes estados a partir de $|0\rangle$, $|1\rangle$,...,$|9\rangle$. Por lo tanto, se necesitan 45 números complejos para especificar un estado de dos fermiones en nuestro sistema.

Para tres fermiones necesitamos tres posiciones distintas y los estados base son de la forma

$$|0\rangle|1\rangle|2\rangle + |1\rangle|2\rangle|0\rangle + |2\rangle|0\rangle|1\rangle - |0\rangle|2\rangle|1\rangle - |2\rangle|1\rangle|0\rangle - |1\rangle|0\rangle|2\rangle.$$

por lo que hay (10 x 9 x 8)/6 = 120 de estos estados en total. De modo que se necesitan 120 números complejos para especificar un estado de tres fermiones. La situación es similar para números más altos de fermiones. Para un par de bosones idénticos, los estados base independientes son de dos tipos; estados como

$$|0\rangle|1\rangle + |1\rangle|0\rangle$$

y estados como

$$|0\rangle|0\rangle$$

(que ahora están permitidos), que dan 10 x 11/2 = 55 en total. Por lo tanto, se necesitan 55 números complejos para nuestros estados de dos bosones. Para tres bosones hay estados base de tres tipos diferentes y se necesitan (10 x 11 x 12)/6 = 220 números complejos. Y así sucesivamente.

Por supuesto que, para mostrar las ideas principales, he estado considerando aquí una situación simplificada. Una descripción más realista requeriría un *continuum* de estados de posición, pero las ideas esenciales son las mismas. Otra pequeña complicación es la presencia de *spin.* Para una partícula de spin 1/2 (necesariamente un fermión) habrá dos estados posibles para cada posición. Caractericémoslos por "↑" (spin " hacia arriba") y "↓," (spin "hacia abajo"). Entonces, para una sola partícula tendremos, en nuestra situación simplificada, veinte estados básicos en lugar de diez:

$$|0\uparrow\rangle, |0\downarrow\rangle, |1\uparrow\rangle, |1\downarrow\rangle, |2\uparrow\rangle, |2\downarrow\rangle, \ldots, |9\uparrow\rangle, |9\downarrow\rangle.$$

Pero, aparte de esto, el procedimiento es igual que antes, de modo que para dos de tales fermiones necesitamos (20 x 19)/2 = 190 números, para tres necesitamos (20 x 19 x 18)/6 = 1140 y así.

En el capítulo I me referí al hecho de que, según la teoría moderna, si se intercambia una partícula del cuerpo de una persona con una partícula similar de uno de los ladrillos de su casa, entonces no sucede nada en absoluto. Si esa partícula fuera un bosón, entonces, como hemos visto, el estado $|\psi\rangle$ quedaría totalmente inafectado. Si la partícula fuera un fermión, entonces el estado $|\psi\rangle$ quedaría reemplazado por $-|\psi\rangle$, que es físicamente idéntico a $|\psi\rangle$. (Podemos remediar este cambio de signo, si lo creemos necesario, sin más que tomar la precaución de dar un giro completo de 360° a una de las dos partículas al hacer el intercambio. Recuérdese que los fermiones cambian de signo bajo una rotación semejante mientras que los estados bosónicos permanecen inalterados.)

La teoría moderna (de alrededor de 1926) dice algo verdaderamente profundo sobre la cuestión de la identidad de fragmentos de material físico: estrictamente hablando, no podemos referirnos a "este electrón particular" o a "ese fotón individual". Afirmar que "el primer electrón está aquí y el segundo allí" es afirmar que el estado tiene la forma $|0\rangle|1\rangle$ que, como hemos visto, no está permitida para un estado fermiónico. Sin embargo, sí es permisible afirmar que "hay un par de electrones: uno aquí y uno allí". Es legítimo referirse al conglomerado de todos los electrones o de todos los protones o de todos los fotones (incluso aunque esto ignore las *interacciones* entre diferentes tipos de partícula). Los electrones individuales proporcionan una aproximación a esta imagen global, como lo hacen los protones individuales o los fotones también individuales. Tal aproximación funciona para la mayoría de los propósitos, pero existen varias circunstancias para las que no lo hace, entre ellas la superconductividad, la superfluidez y el comportamiento del láser.

La imagen del mundo físico que nos ha presentado la mecánica cuántica no es en absoluto la imagen a la que nos había acostumbrado la física clásica. Pero sujétense el sombrero: todavía hay cosas más extrañas en el mundo cuántico.

LA "PARADOJA" DE EINSTEIN, PODOLSKY Y ROSEN

Como se mencionó al principio de este capítulo, algunas de las ideas de Albert Einstein fueron fundamentales para el desarrollo de la teoría cuántica. Recuérdese que fue él quien propuso por primera vez (en 1905) el concepto del "fotón" —el cuanto de campo electromagnético—, a partir del cual se desarrolló la idea de la dualidad onda-corpúsculo. (El concepto de "bosón" fue también en parte suyo, como lo fueron muchas otras aportaciones centrales a la teoría.) Pese a todo, Einstein nunca pudo aceptar que esa teoría que se desarrolló a partir de sus ideas pudiera llegar a ser algo más que una mera descripción provisional del mundo físico. Su aversión frente al aspecto probabilístico de la teoría fue bien conocida desde el principio y quedó recogida en su contestación a una de las cartas de Max Born en 1926 (citada en Pais, 1982, p. 443):

> La mecánica cuántica es algo muy serio. Pero una voz interior me dice que, de todos modos, no es ése el camino. La teoría dice mucho, pero en realidad no nos acerca demasiado al secreto del Viejo. En todo caso estoy convencido de que Él no juega a los dados.

Parece, sin embargo, que lo que molestaba a Einstein, más incluso que ese indeterminismo físico, era una aparente *falta de objetividad* en la manera empleada para describir la teoría cuántica. En mi exposición sobre ésta he procurado destacar que la descripción del mundo que proporciona es bastante objetiva, aunque a veces muy extraña y contraria a la intuición. Por el contrario, Bohr parece haber considerado el estado cuántico de un sistema (entre medidas) como carente de genuina realidad física, como un resumen de *nuestro conocimiento* acerca de dicho sistema. Pero ¿no podrían los distintos observadores tener un conocimiento diferente de un sistema, de tal modo que la función de onda parecería ser algo esencialmente *subjetivo* o "sólo presente en la mente de los físicos"?

No deberíamos permitir que nuestra maravillosamente precisa imagen física del mundo, desarrollada durante muchos siglos se evaporara. Para ello Bohr necesitaba haber considerado que el mundo en el *nivel clásico* tenía una realidad objetiva, pero que no habría "realidad" en los estados de nivel *cuántico* que parecen subyacer en todo.

Semejante imagen era inadmisible para Einstein, quien creía que debe haber un mundo físico objetivo, incluso en las escalas minúsculas de los fenómenos cuánticos. En sus numerosas discusiones con Bohr intentó (aunque sin éxito) demostrar que había contradicciones inherentes a la imagen cuántica de las cosas, y que debe haber una estructura aún más profunda debajo de la teoría cuántica, probablemente más afín a la imagen que nos había presentado la física clásica. Tal vez bajo los comportamientos probabilistas de la mecánica cuántica subyaciera la acción estadística de ingredientes o "partes" más pequeñas del sistema de las que no tenemos conocimiento directo. Los seguidores de Einstein, en particular David Bohm, desarrollaron el punto de vista de las "variables ocultas", según el cual hay alguna realidad definida, pero los parámetros que precisa el sistema no nos son directamente accesibles, y las probabilidades cuánticas surgen porque no se pueden conocer los valores de esos parámetros antes de la medida.

¿Puede una teoría de variables ocultas ser consistente con todos los hechos observacionales de la física cuántica? La respuesta es sí, pero sólo si la teoría es esencialmente *no local* en el sentido de que los parámetros ocultos sean capaces de afectar instantáneamente a regiones del sistema arbitrariamente lejanas. *Esto* no le hubiera gustado a Einstein, debido, en particular a las dificultades que surgen para la relatividad especial. Las consideraré más adelante.

La teoría de variables ocultas más afortunada es la que se conoce como modelo de De Broglie-Bohm (De Broglie, 1956; Bohm, 1952). No discutiré aquí estos modelos, puesto que mi propósito en este capítulo es dar sólo una visión general de la teoría cuántica estándar, y no detallar las propuestas rivales. Si queremos objetividad física y estamos dispuestos a prescindir del determinismo, la teoría estándar será suficiente. Consideremos simplemente que el vector de estado corresponde a la "realidad" y que evoluciona normalmente según el procedimiento determinista continuo **U**, pero saltando de vez en cuando y de forma extraña según **R**, cuando quiera que un efecto se amplifica hasta el nivel clásico. Sin embargo, persiste el problema de la no localidad y las dificultades aparentes con la relatividad. Echémosles una mirada.

Supongamos que tenemos un sistema físico que consta de dos subsistemas **A** y **B**. Consideremos que **A** y **B** son dos partículas diferentes. Supongamos que las dos opciones (ortonormales) para el estado **A** son $|\alpha\rangle$ y $|\rho\rangle$, mientras que el estado de **B** podría ser $|\beta\rangle$ o $|\sigma\rangle$. Como hemos visto antes, el estado general combinado no será simplemente un producto ("y") *de* un estado de **A** por un estado de **B**, sino una superposición ("más") de tales productos. (Diremos entonces que **A** y **B** están *correlacionados.)* Consideremos que el estado del sistema es

$$|\alpha\rangle|\beta\rangle+|\rho\rangle|\sigma\rangle$$

Realicemos ahora una medida *sí/no* sobre A que distinga $|\alpha\rangle$ ("sí") de $|\rho\rangle$ ("no"). ¿Qué pasa con **B**? Si la medida da "sí", entonces el estado resultante debe ser

$$|\alpha\rangle|\beta\rangle$$

mientras que si da "no", entonces es

$$|\rho\rangle|\sigma\rangle$$

Por consiguiente, nuestra medida de **A** provoca que el estado de **B** salte: a $|\beta\rangle$ en el caso de una respuesta "sí"; y a $|\sigma\rangle$, en el de una respuesta "no". La partícula **B** no necesita estar cerca de **A**; podría estar a años luz de distancia. Pero **B** salta en el momento de medir **A**.

A ver, un momento —puede estar diciendo el lector—. ¿Qué es este pretendido *salto?* ¿Por qué no son las cosas de esta otra forma? Imaginemos una caja de la que sabemos que contiene una bola blanca y una bola negra. Supongamos que se sacan las bolas **U**. Sin mirarlas, se llevan a dos rincones opuestos de la habitación. Entonces, si se examina una bola y resulta ser blanca (el análogo de $|\alpha\rangle$ arriba), la otra debe ser negra (el análogo de $|\beta\rangle$). Si, por el contrario, la primera resulta ser la bola negra ($|\rho\rangle$*)*, entonces, el estado incierto de la segunda bola salta a "blanco, con certeza" ($|\sigma\rangle$). Nadie en su sano juicio, insistirá el lector, atribuirá el cambio repentino del estado "incierto" de la segunda bola —hasta ser "negra, con certeza" o "blanca, con certeza"— a alguna misteriosa "influencia" no local que se propagase instantáneamente desde la primera bola en el mismo momento en que ésta es examinada.

Pero la naturaleza es en realidad extraordinaria. En lo que antecede podríamos imaginar que el *sistema* "sabía" ya, por ejemplo, que el estado de **B** era $|\beta\rangle$ y el de **A** era $|\alpha\rangle$ (o que el de **B** era $|\sigma\rangle$ y el de A era $|\rho\rangle$) antes de que se realizase la medida en **A**, y el *experimentador* era el que no lo sabía. Al encontrar que **A** está en el estado $|\alpha\rangle$, éste simplemente *infiere* que **B** está en $|\beta\rangle$. Ese sería un punto de vista "clásico" —semejante al de una teoría local de variables ocultas— y no habría lugar a ningún "salto" *físico.* (Todo estaría en la mente del experimentador.) Según esta idea, cada parte del sistema "sabe" por adelantado los resultados d cualquier experimento que se pudiera realizar sobre ella. Las probabilidades aparecen debido únicamente a la falta de conocimiento por parte del experimentador. Pero resulta que este punto de vista *no funcionará* como explicación de las enigmáticas probabilidades aparentemente no locales que hay en la teoría cuántica.

Para ver esto consideraremos una situación como la anterior pero en la que la *elección de la medida* en el sistema **A** no se decide hasta que **A** y **B** están muy separados. El comportamiento de **B** parece estar influido instantáneamente por esta misma elección. Este tipo EPR de experimento mental aparentemente paradójico se debe a Albert Einstein, Boris Podolsky y Nathan Rosen (1935).

Daré una variante propuesta por David Bohm (1951). El hecho de que ninguna descripción local "realista" (digamos de variables ocultas, o de "tipo clásico") pueda dar las probabilidades cuánticas correctas se sigue de un famoso teorema debido a John S. Bell. (Véase Bell, 1987; Rae, 1986, Squires, 1986.) Supongamos que se producen dos partículas de spin 1/2 —que llamaré *electrón* y *positrón* (esto es, un *antielectrón)*— en la desintegración de una simple partícula de spin cero en algún punto central, y que las dos se alejan en direcciones exactamente opuestas (fig. VI.30).

Por la conservación del momento angular, los spines del electrón y el positrón deben sumar cero porque ése era el momento angular de la partícula inicial. Esto tiene como consecuencia que cuando medimos el spin del electrón en alguna dirección, cualquiera que sea ésta, el positrón tiene un spin en la dirección *opuesta.* Las dos partículas podrán estar a kilómetros o incluso años luz de distancia, pero esa misma *elección* de la medida en una partícula puede haber fijado *instantáneamente* el eje de giro de la otra.

Veamos cómo el formalismo cuántico conduce a esta conclusión. Representamos el estado combinado de momento angular nulo de las dos partículas mediante el vector de estado $|Q\rangle$, y encontramos una relación como

$$|Q\rangle=|E\uparrow\rangle|P\downarrow\rangle-|E\downarrow\rangle|P\uparrow\rangle,$$

en donde E designa al electrón y P al positrón.

Aquí se han descrito las cosas en términos de las direcciones *arriba/abajo* del spin. El estado total es una superposición lineal del electrón con spin *hacia arriba* y el positrón con spin *hacia abajo,* y del electrón con spin *hacia abajo* y el positrón con spin *hacia arriba.* Si medimos el spin del electrón en la dirección *arriba/abajo* y encontramos que en realidad es arriba, entonces debemos saltar al estado $|E\uparrow\rangle|P\downarrow\rangle$, de modo que el estado de spin del positrón debe ser abajo. Si, por el contrario, encontramos que el spin del electrón es abajo, entonces el estado salta a $|E\downarrow\rangle|P\uparrow\rangle$, de modo que el spin del positrón es arriba.

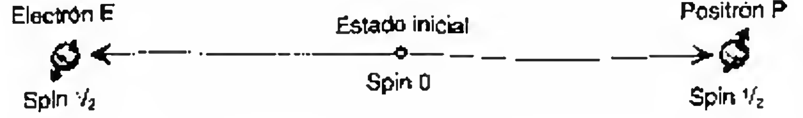


FIGURA VI.30. *Una partícula de spin cero se desintegra en dos partículas de spin 1/2, un electrón E* y *un positrón P. La medida del spin de una de las partículas de spin ½ fija en apariencia* instantáneamente *el estado de spin de la otra.*

Supongamos ahora que hemos elegido algún otro par de direcciones opuestas, por ejemplo derecha e izquierda, en donde

$$|E\rightarrow\rangle=|E\uparrow\rangle+|E\downarrow\rangle \qquad |P\rightarrow\rangle=|P\uparrow\rangle+|P\downarrow\rangle$$

$$|E\leftarrow\rangle=|E\uparrow\rangle-|E\downarrow\rangle \qquad |P\leftarrow\rangle=|P\uparrow\rangle-|P\downarrow\rangle$$

Entonces encontramos (pueden comprobar el álgebra, si quieren):

$$|E\rightarrow\rangle|P\leftarrow\rangle-|E\leftarrow\rangle|P\rightarrow\rangle$$

$$=(|E\uparrow\rangle+|E\downarrow\rangle)(|P\uparrow\rangle-|P\downarrow\rangle)-(|E\uparrow\rangle-|E\downarrow\rangle)(|P\uparrow\rangle+|P\downarrow\rangle)$$

$$=|E\uparrow\rangle|P\uparrow\rangle+|E\downarrow\rangle|P\uparrow\rangle-|E\uparrow\rangle|P\downarrow\rangle-|E\downarrow\rangle|P\downarrow\rangle-|E\uparrow\rangle|P\uparrow\rangle+|E\downarrow\rangle|P\uparrow\rangle-|E\uparrow\rangle|P\downarrow\rangle+|E\downarrow\rangle|P\downarrow\rangle$$

$$=-2(|E\uparrow\rangle|P\downarrow\rangle-|E\downarrow\rangle|P\uparrow\rangle)$$

$$=-2|Q\rangle$$

que, aparte del factor -2, que no es importante, es el mismo estado que el de partida.

Por lo tanto, nuestro estado original puede considerarse también como una superposición lineal del electrón con spin hacia la derecha y el positrón hacia la izquierda, y el del electrón con spin hacia la izquierda y el positrón hacia la derecha. Esta expresión es útil si decidimos medir el spin del electrón en una dirección *derecha/izquierda* en lugar de *arriba/abajo.* Si encontramos que el spin está hacia la derecha, entonces el estado salta a $|E\rightarrow\rangle|P\leftarrow\rangle$, de modo que el positrón tiene spin *hacia la izquierda.* Si, por el contrario, encontramos que el electrón tiene spin *hacia la izquierda,* entonces el estado salta a $|E\leftarrow\rangle|P\rightarrow\rangle$ de modo que el positrón tiene spin *hacia la derecha.* Si hubiéramos decidido medir el spin del electrón en cualquier otra dirección, la historia sería exactamente análoga: el estado de spin del positrón saltaría instantáneamente para estar en esa dirección o en la opuesta, dependiendo del resultado de la medición sobre el electrón.

¿Por qué no podemos hacer un modelo de los spines de nuestros electrón y positrón de la misma forma que en el ejemplo anterior de una bola negra y una bola blanca sacadas de una caja? Hagámoslo de una forma totalmente general.

En lugar de tener una bola negra y una bola blanca podríamos tener dos elementos mecánicos E y P que estuvieran inicialmente unidos y que luego separaremos en direcciones opuestas. Supongamos que cada uno de estos E y P puede dar lugar a una respuesta "sí" o "no" a una medida del spin en cualquier dirección dada. Esta respuesta podría estar completamente determinada por la constitución mecánica para cada elección de la dirección —o quizá el mecanismo produjera sólo respuestas probabilísticas, determinadas por la constitución mecánica— pero donde suponemos que, tras la separación, *cada uno de los* E *y* P *se comporta de forma totalmente independiente del otro.*

Disponemos de medidores de spin en cada lado, uno que mide el spin de E y otro que mide el de P. Supongamos que hay tres posiciones para la dirección del spin en cada medidor, digamos A, B y C para el medidor E, y A', B' y C' para el medidor P. Las direcciones A', B' y C' serán paralelas respectivamente a las A, B y C, y consideraremos que A, B y C están en el mismo plano y formando ángulos iguales, es decir, a 120° una de otra. (Véase fig. VI.31.)

El experimento se repite muchas veces, con valores diferentes de estas posiciones en cada lado. A veces el medidor E registrará "sí" (es decir, el spin *está* en la dirección medida: A o B o C) y a veces registrará "no" (spin en la dirección opuesta). Análogamente, el medidor P registrará a veces "sí" y a veces "no". Tomemos ahora nota de dos propiedades que deben tener las verdaderas probabilidades *cuánticas:*

1) Si las posiciones en los dos lados son *iguales* (esto es, A y A', etc.), los resultados de las dos medidas *discrepan* (es decir, el medidor E registra "sí" siempre que el medidor P registre "no", y "no" siempre que P dé "sí").

2) Si los diales de las posiciones se giran y fijan *aleatoriamente,* independientes uno de otro, los dos medidores tienen tanto la *misma probabilidad de coincidir como la de discrepar.*

Podemos ver fácilmente que las propiedades *1)* y *2)* se siguen directamente de las reglas de probabilidad cuántica que hemos dado antes. Supongamos que el medidor E actúa primero. El medidor P encuentra entonces una partícula cuyo estado de spin es opuesto al medido por el medidor E, de modo que la propiedad *1)* se sigue inmediatamente. Si queremos obtener la propiedad *2)* notemos que, para direcciones medidas que forman un ángulo de 120°, si el medidor E da "sí" entonces la dirección de P está a 60° del estado de spin sobre el que actúa y si es ¨no¨ entonces esta a 120° del estado de spin.

En consecuencia, existe una probabilidad 3/4 = 1/2(1+cos 60°) de que las medidas coincidan y una probabilidad 1/4 = ½(1+cos 120º) de que discrepen. Por lo tanto, la probabilidad promediada para las tres posiciones de P, si E da ¨si¨, es 1/3(0+3/4+3/4)=1/2 para P dando ¨si¨, y 1/3(1+1/4+1/4)=1/2 para P dando ¨no¨, es decir igualmente probable el acuerdo que la discrepancia. Y análogamente si da ¨no¨. Esta es de hecho la propiedad *2).*

Es un hecho notable que *1)* y *2)* sean *inconsistentes* con cualquier modelo realista local. Supongamos que tuviéramos tal modelo. La máquina E debe ser preparada para cada una de las medidas posibles A, B o C. Si estuviera preparada sólo para dar una respuesta *probabilística,*

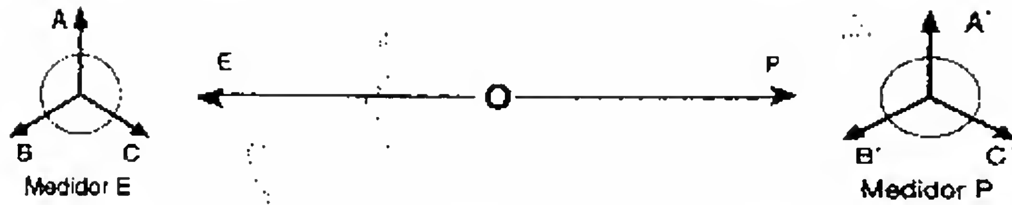


FIGURA VI. 31. *Versión sencilla de David Mermin de la paradoja EPR y el teorema de Bell, que pone de relieve una contradicción entre una visión realista local de la naturaleza y los resultados de la teoría cuántica. El medidor* **E** *y el medidor* **P**, *independientes uno de otro, tienen tres posiciones para la dirección en que pueden medir los spines de sus respectivas partículas.*

entonces no podría *asegurarse* que la máquina P estuviera en desacuerdo con ella, para A', B' y C', respectivamente, según afirma *1)*. En realidad, *ambas* máquinas deben tener preparadas por adelantado sus respuestas a cada una de las tres posibles medidas.

Supongamos, por ejemplo, que tales respuestas van a ser "sí", "sí", "sí" respectivamente, para A, B, C. La partícula del lado derecho debe entonces estar preparada para dar "no", "no", "no" para las tres posiciones correspondientes, en el lado derecho. Si en lugar de esto las respuestas preparadas en el lado izquierdo van a ser "sí", "sí", "no", entonces las respuestas del lado derecho deben ser "no", "no", "sí". Todos los demás casos serían esencialmente similares a éstos.

Veamos ahora si esto puede ser compatible con *2). Las* asignaciones "sí", "sí", "sí" / "no", "no", "no" no son muy prometedoras porque eso da 9 casos de discrepancia y O casos de acuerdo en todos los emparejamientos posibles A/A', A/B', A/C, B/A', etc. ¿Qué sucede con "sí", "sí", "no"/"no","no","sí" y similares? Éstos dan 5 casos de discrepancia y 4 de acuerdo. (Para comprobarlo no hay más que contarlos: S/N, S/N, S/S, S/N, S/N, S/S, N/N, N/N, N/S, 5 de los cuales discrepan y 4 coinciden.) Lo cual está mucho más cerca de lo que se necesita para *2),* pero *no* es suficiente, ya que necesitamos tantas coincidencias como discrepancias. Cualquier otro par de asignaciones consistentes con *1)* darían de nuevo 5 frente a 4 (excepto para "no", "no", "no" / "sí", "sí", "sí" que es peor, porque da, otra vez, 9 a 0). *No* existe conjunto de respuestas preparadas que pueda dar lugar a las probabilidades cuánticas. *Los modelos realistas locales quedan descartados.*[14]

---

[14] Este es un resultado tan famoso e importante que vale la pena dar otra versión de él. Supongamos que hay sólo *dos* posiciones para el medidor E, arriba [↑] y derecha [→], y dos para el medidor P, 45° hacia arriba y a la derecha [%] y 45° hacia abajo y a la derecha [Â]. Considérese que las posiciones *reales* son [→] y [%], para los medidores E y P, respectivamente. Entonces la probabilidad de que los medidores E y P coincidan es 1/2(1 + cos135°) = 0.146..., que es un poco menos del 15%. Una larga serie de experimentos con estas posiciones dadas, por ejemplo,

E:SNNSNSSSNSSNNSNNNNSSN...
P:NSSNNNSNSNNSSNSSNSNNS...

dará precisamente algo menos de 15% de coincidencias. Supongamos ahora que las medidas-P no están influidas por la posición-E —de modo que *si* la posición-E hubiera sido [↑] en lugar de [→], entonces la serie de resultados-P habría sido exactamente la misma— y puesto que el ángulo entre [↑] y [%] es el mismo que el ángulo entre [→] y [%], habría otra vez algo menos de 15% de coincidencias entre las medidas-P y las nuevas medidas-E, digamos E'. Por otro lado, si la posición-E hubiera sido [→], como antes, pero la posición-P fuera [Â] en lugar de [%], entonces la serie de resultados-E habría sido la misma que antes pero los nuevos resultados-P, digamos P', estarían por debajo del 15% de coincidencias con los resultados-E originales. Se sigue que no

EXPERIMENTOS CON FOTONES: ¿UN PROBLEMA PARA LA RELATIVIDAD?

Debemos preguntarnos si la experiencia corrobora las expectativas cuánticas. El ejemplo anterior es un experimento hipotético, pero experimentos similares se *han realizado* utilizando las polarizaciones de pares de *fotones* en lugar del spin de partículas con masa y de spin y. Aparte de esta diferencia, esos experimentos son, en sus aspectos esenciales, iguales al descrito, excepto que los ángulos que nos interesan (puesto que los fotones tienen spin uno en lugar de 1/2) serán precisamente los ángulos mitad del experimento con partículas de spin 1/2.

La polarización de los pares de fotones se ha medido en varias combinaciones de direcciones y los resultados están de acuerdo con las predicciones de la teoría cuántica, e inconsistentes con cualquier modelo realista local.

Los resultados más precisos y convincentes entre los obtenidos hasta la fecha son los de Alain Aspect (1986) y sus colegas en París.[15] Los experimentos de Aspect tienen otra característica interesante. Las "decisiones" sobre la forma de medición de las polarizaciones de los fotones se tomaban solamente después de que los fotones estuvieran en vuelo. Por lo tanto, si pensamos que alguna "influencia" no local está viajando desde un detector de fotones al fotón del lado opuesto, indicando la dirección en la que intenta medir la dirección de polarización del fotón que se aproxima, entonces vemos que esta "influencia" debe viajar más rápida que la luz. Cualquier tipo de descripción realista del mundo cuántico que sea consistente con los hechos debe ser aparentemente *no causal,* en el sentido de que los efectos deben poder viajar mas rápidos que la luz. Pero vimos en el último capítulo que, mientras sea válida la relatividad general, el envío de señales más rápidas que la luz conduce a absurdos (y entra en conflicto con nuestras nociones de "libre albedrío", etc). Esto es ciertamente verdadero, pero las "influencias" no locales que aparecen en experimentos tipo EPR no son de las que se pueden utilizar para enviar mensajes, como se puede ver por la misma razón de que, si lo fueran, conducirían a absurdos semejantes. (Una demostración detallada de que estas "influencias" no pueden utilizarse para enviar mensajes ha sido llevada a cabo por Ghirardi, Rimini y Weber, 1980.) No sirve de nada que se nos diga que un fotón está polarizado "o vertical u horizontalmente" (como opuesto, digamos, "o a 60° o a 150°") hasta que se nos informe de cuál de las dos opciones es la verdadera. Es el primer elemento de "información" (es decir, las *direcciones* de polarización posibles) el que llega más rápido que la luz (instantáneamente), mientras que el conocimiento sobre en *cuál* de estas direcciones deberá estar polarizado efectivamente llega más lentamente, por vía de una señal ordinaria que comunica el *resultado* de la primera medida de polarización.

Aunque los experimentos de tipo EPR no entran en conflicto, en el sentido ordinario del envío de señales, con la *causalidad* de la relatividad existe un conflicto indudable con el *espíritu* de la

---

podría haber más de 45% (15% + 15% + 15%) de coincidencias entre la medida-P' [Ê] y la medida-E' [↑] si hubieran sido *éstas* las posiciones reales. Pero el ángulo entre [Ê] y [↑] es 135° y no 45°, de modo que la probabilidad de coincidencia *debería* ser justo algo mayor del 85%, no el 45%. Esto es una contradicción que muestra que la hipótesis de que la elección de la medida a hacer en E no puede influir en los resultados de P (y *viceversa)* debe ser falsa! Estoy en deuda con David Mermin por este ejemplo. La versión dada en el texto principal está tomada de su artículo (Mermin, 1985).

[15] Los primeros resultados fueron debidos a Freedman y Clauser (1972) basados en ideas sugeridas por Clauser, Home, Shimony y Holt (1969). Existe todavía un punto de discusión, en estos experimentos, debido al hecho de que los detectores de fotones que se utilizan están muy lejos de alcanzar una eficiencia del 100%, de modo que sólo una fracción relativamente pequeña de los fotones emitidos son realmente detectados. Sin embargo, el acuerdo con la teoría cuántica es tan perfecto, con estos detectores relativamente ineficientes, que es difícil ver como la mejora de los detectores vaya a producir un acuerdo *peor* con la teoría!

relatividad en nuestra imagen de la "realidad física". Veamos cómo la visión *realista* del vector de estado se aplica al experimento de tipo EPR anterior (con fotones) A medida que se separan los dos fotones, el vector de estado describe la situación como un *par* de fotones que actúa como una sola unidad. Ninguno de los fotones por separado tiene un estado objetivo: el estado cuántico se aplica sólo a los dos en conjunto. Ninguno de los fotones tiene individualmente, una dirección de polarización: la polarización es una cualidad combinada de ambos fotones juntos. Cuando se mide la polarización de uno de estos dos fotones, el vector de estado *salta* de modo que el fotón no medido *tiene* ahora una polarización definida. Cuando *dicha* polarización del fotón es medida posteriormente, los valores de la probabilidad se obtienen correctamente aplicando las reglas cuánticas usuales a su estado de polarización. Esta manera de considerar la situación proporciona las respuestas correctas; es, en efecto, el modo en que aplicamos ordinariamente la mecánica cuántica. Pero es una visión esencialmente no-relativista; las dos medidas de polarización están separadas con lo que se conoce como un *intervalo de tipo espacio,* lo que significa que cada una está fuera del cono de luz de la otra, como los puntos *R* y Q en la fig. V.21. La pregunta de cuál de estas dos medidas ocurrió *primero* no es físicamente significativa sino que depende del estado de movimiento del "observador" (véase fig. VI.32). Si el "observador" se mueve suficientemente rápido hacia la derecha, entonces él considera que la medida en el lado derecho ha ocurrido primero; y si se mueve hacia la izquierda, entonces es la medida en el lado izquierdo. Pero si consideramos que el fotón del lado derecho ha sido medido primero obtenemos una imagen de la realidad física completamente diferente de la obtenida si consideramos que el fotón del lado izquierdo se ha medido primero. (Es una medida diferente la que causa el "salto" no local.) Hay un conflicto esencial entre nuestra imagen espacio-temporal de la realidad física —incluso la mecánica-cuántica correctamente no local— y la relatividad especial. Este es un serio enigma, que los "realistas cuánticos" no han podido resolver adecuadamente *(cfr.* Aharonov y Albert, 1981). Tendré necesidad de volver al tema más adelante.

La ecuación de Schrödinger y la ecuación de Dirac

Antes, en este mismo capítulo, me he referido a la ecuación de Schrödinger que es una ecuación determinista perfectamente definida, similar en

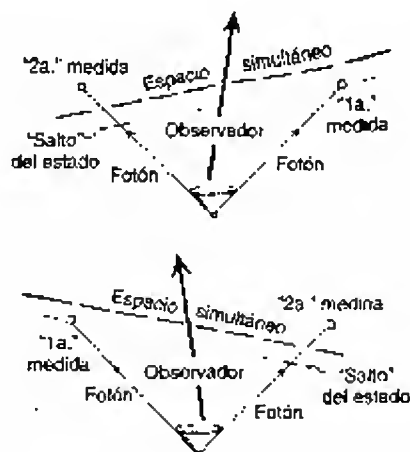


FIGURA VI.32. *Dos observadores diferentes forman imágenes de la "realidad" mutuamente inconsistentes en un experimento EPR en el que dos fotones son emitidos en direcciones opuestas a partir de un estado de spin 0. El observador que se mueve hacia la derecha juzga que*

la parte izquierda del estado salta antes *de que sea medido, estando causado el salto por la medida en la derecha. El observador que se mueve hacia la izquierda tiene la idea contraria.*

muchos aspectos a las ecuaciones de la física clásica. Las reglas dicen que mientras no se hagan "medidas" (u "observaciones") sobre un sistema cuántico, la ecuación de Schrödinger será válida. El lector puede ser testigo de su forma real:

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle$$

Recordemos que $\hbar$ es una versión de Dirac de la constante de Planck *($h/2\pi$)* (e i = $\sqrt{-1}$) y que el operador $\partial/\partial t$ (derivada parcial respecto al tiempo) actuando sobre $|\psi\rangle$ significa simplemente el *ritmo de cambio* de $|\psi\rangle$ con respecto al tiempo. La ecuación de Schrödinger establece que *"$H|\psi\rangle$"* describe cómo evoluciona $|\psi\rangle$.

Pero ¿qué es *"H"* Es la *función hamiltoniana* que consideramos en *el* capítulo anterior pero con una diferencia fundamental. Recuérdese que el hamiltoniano clásico es la expresión para la *energía total* en términos de las diversas coordenadas de posición q_i y coordenadas de momento p_i, *para* todos los objetos físicos en el sistema. Para obtener el hamiltoniano *cuántico* tomamos la misma expresión, pero cada vez que aparezca el momento p_i, lo sustituimos por un múltiplo del *operador diferencial* "derivada parcial respecto a q_i". Concretamente, reemplazamos p_i por -$i\hbar\,\partial/\partial q_i$,. Nuestro hamiltoniano cuántico *H* se convierte entonces en una (frecuentemente complicada) *operación* matemática que incluye derivadas y multiplicaciones, etc., y ya no un simple número. Esto parece una especie de abracadabra. Sin embargo, no es sólo un conjuro matemático: es auténtica *magia* en acción. (Hay un poco de "arte" en la aplicación de este proceso de generar un hamiltoniano cuántico a partir de uno clásico, pero es curioso, en vista de su extravagante naturaleza, lo poco que parecen importar las ambigüedades inherentes al procedimiento.)

Algo importante que señalar acerca de la ecuación de Schrödinger (cualquiera que sea *H)* es que es *lineal,* es decir, si $|\psi\rangle$ y $|\varphi\rangle$ satisfacen ambos la ecuación, entonces también lo hace $|\psi\rangle + |\varphi\rangle$ —o, de hecho, cualquier combinación $w\,|\psi\rangle + z\,|\varphi\rangle$, donde w y z son números complejos fijos. Por lo tanto, una superposición lineal compleja se sigue manteniendo indefinidamente mediante la ecuación de Schrödinger. Una superposición lineal (compleja) de dos estados alternativos posibles no puede "*des*-superponerse" simplemente por la acción de **U**. Por esta razón es necesaria la acción de **R**, como un procedimiento *independiente* de **U**, para que finalmente sobreviva sólo *una* de las opciones.

Al igual que el formalismo hamiltoniano para la física clásica, la ecuación de Schrödinger no es tanto una ecuación concreta como un marco para las ecuaciones mecánico-cuánticas en general. Una vez que se ha obtenido el hamiltoniano cuántico apropiado, la evolución temporal del estado según la ecuación de Schrödinger tiene lugar como si $|\psi\rangle$ fuera un campo clásico sujeto a alguna ecuación clásica de campos como las de Maxwell. De hecho, si $|\psi\rangle$ describe el estado de un simple *fotón,* entonces resulta que la ecuación de Schrödinger se *convierte* realmente en las ecuaciones de Maxwell. La ecuación para un simple fotón es exactamente la misma que la ecuación[*] para un campo electromagnético completo. Este hecho es responsable de que el

[*] Sin embargo, hay una diferencia importante en el tipo de *solución* a la ecuación que es admisible. Los campos de Maxwell clásicos son necesariamente *reales* mientras que los estados del fotón son *complejos.* Existe también una condición llamada de "frecuencia positiva" que debe satisfacer el estado del fotón.

comportamiento de los *fotones individuales* que atisbamos antes sea parecido a las ondas del campo de Maxwell y la polarización. Como un ejemplo distinto, si $|\psi\rangle$ describe el estado de un simple *electrón,* entonces la ecuación de Schrödinger se transforma en la famosa ecuación de ondas de Dirac para el electrón, descubierta en 1928 después de que Dirac hubiera puesto mucha originalidad e intuición adicional.

De hecho, la ecuación de Dirac para el electrón debe colocarse al lado de las ecuaciones de Maxwell y de Einstein, como una de las grandes ecuaciones de campos de la física. Dar aquí una idea correcta de ella me obligaría a introducir ideas matemáticas que nos distraerían demasiado. Baste decir que en la ecuación de Dirac $|\psi\rangle$ tiene la curiosa propiedad "fermiónica" $|\psi\rangle \to -|\psi\rangle$ bajo la rotación de 360° que consideramos antes. Las ecuaciones de Dirac y Maxwell juntas constituyen los ingredientes básicos de la electrodinámica cuántica, la más acertada de las teorías cuánticas de campos. Consideraremos esto en breve.

LA TEORIA CUÁNTICA DE CAMPOS

La materia conocida como "teoría cuántica de campos" ha aparecido como una unión de las ideas procedentes de la relatividad especial y la mecánica cuántica. Difiere de la mecánica cuántica estándar (es decir, no relativista) en que el número de partículas, de cualquier tipo, no necesita ser constante. Cada tipo de partícula tiene su *antipartícula* (a veces, como en el caso de los fotones, la misma que la partícula original). Una partícula con masa y su antipartícula pueden aniquilarse para dar energía, y un par partícula-antipartícula puede crearse a partir de la energía. En realidad, el número de partículas no necesita estar definido; se permiten superposiciones lineales de estados con diferentes números de partículas. La teoría cuántica de campos suprema es la "electrodinámica cuántica" —básicamente la teoría de electrones y fotones—. Esta teoría es notable por la precisión de sus predicciones *(v.g.* el valor preciso del momento magnético del electrón, citado en el anterior capítulo). Sin embargo, es una teoría más bien desordenada —y no consistente globalmente— debido a que inicialmente da respuestas "infinitas" sin sentido. Éstas deben ser eliminadas mediante un proceso conocido como "renormalización". No todas las teorías cuánticas de campos son susceptibles de renormalización, y es difícil calcular con ellas incluso cuando lo son. Una aproximación popular a la teoría cuántica de campos es vía "integrales de camino", que supone la formación de superposiciones lineales cuánticas no sólo de diferentes estados de partículas (como con las funciones de onda ordinarias) sino de historias completas espacio-temporales de comportamiento físico (véase Feynman, 1985, para una presentación popular). Sin embargo, esta aproximación tiene infinitos adicionales propios de ella, y sólo se puede entender vía la introducción de diversos "trucos matemáticos". A pesar de la potencia indudable y la impresionante precisión de la teoría cuántica de campos (en aquellos pocos casos en los que la teoría se puede llevar hasta el final), nos quedamos con una sensación de que se necesita una comprensión más profunda antes de poder estar seguros de cualquier "imagen de la realidad física" a la que parezca conducir.[16]

Debería dejar claro que la compatibilidad entre la teoría cuántica y la relatividad especial que proporciona la teoría cuántica de campos es sólo *parcial* — sólo afecta a **U** — y es sobre todo de naturaleza matemáticamente formal. La dificultad de una interpretación relativísticamente consistente de los "saltos cuánticos" que ocurren con **R**, la que nos dejaron los experimentos de

[16] La teoría cuántica de campos parece ofrecer alguna perspectiva para la no computabilidad *(cfr.* Komar, 1964).

tipo EPR, no es ni siquiera esbozada por la teoría cuántica de campos. Tampoco hay todavía ninguna teoría cuántica de campo gravitatorio consistente o creíble. Sugeriré, en el capítulo VIII, que estos temas pueden no estar completamente disociados.

EL GATO DE SCHRÖDINGER

Volvamos finalmente a un tema que nos ha perseguido desde los inicios de nuestras descripciones. ¿Por qué no vemos superposiciones lineales de objetos a escala clásica, como bolas de cricket en dos lugares a la vez? ¿Qué es lo que hace que ciertas disposiciones de átomos constituyan "dispositivos de medida", de modo que el procedimiento **R** parezca reemplazar a **U**? Ciertamente, cualquier pieza de un aparato de medida es ella misma parte del mundo físico, construida a partir de aquellos constituyentes cuánticos para cuyo examen del comportamiento ha sido diseñada. ¿Por qué no tratar el aparato de medida *junto* con el sistema físico examinado como un *sistema cuántico combinado.* Ninguna misteriosa medida "externa" está ahora involucrada. El sistema combinado debería evolucionar simplemente según **U**. ¿Pero lo hace? La acción de **U** sobre el sistema combinado *es* completamente determinista, sin lugar para las incertidumbres probabilistas de tipo **R** implicadas en la "medición" u "observación" que el sistema combinado está realizando sobre sí mismo. Hay aquí una aparente contradicción, hecha especialmente gráfica en un famoso experimento mental introducido por Erwin Schrödinger (1935): *la paradoja del gato de Schrödinger.* Imaginemos una habitación cerrada, construida de forma tan perfecta que ninguna influencia física puede atravesar sus paredes ni hacia dentro ni hacia afuera. Imaginemos que dentro de la habitación hay un gato y también un dispositivo que puede ser disparado mediante algún suceso cuántico. Si tiene lugar dicho suceso, entonces el dispositivo rompe una ampolla que contiene cianuro y el gato muere. Si el suceso no tiene lugar, el gato continúa vivo. En la versión original de Schrödinger el suceso cuántico era la desintegración de un átomo radioactivo. Permítaseme modificarla ligeramente y supongamos que el suceso cuántico es el disparo de una fotocélula por un fotón, fotón que ha sido emitido por alguna fuente luminosa en cierto estado predeterminado, y luego reflejado en un espejo semirreflectante (véase fig. VI.33) La reflexión en el espejo desdobla la función de onda del fotón en dos partes separadas, una de las cuales se refleja y la otra se transmite a través del espejo. La parte reflejada de la función de onda del fotón *se* localiza sobre la fotocélula, de modo que si el fotón *es* registrado por la fotocélula ello significa que ha sido *reflejado.* En tal caso se libera el cianuro y el gato muere. Si la fotocélula no registra nada, el fotón fue *transmitido* a través del espejo semirreflectante hasta la pared que hay detrás, y el gato se salva.

Desde el punto de vista (algo peligroso) *de* un observador en el *interior* de la habitación, ésta sería la descripción de lo que allí ocurría. (Hubiéramos hecho mejor en proporcionar a este observador un traje protector adecuado.) *O bien* se considera que el fotón ha sido reflejado, porque se "observa" que la fotocélula ha registrado y el gato ha muerto, *o bien* se considera que el fotón ha sido transmitido, porque se "observa" que la fotocélula *no* ha registrado nada y el gato está vivo. O lo uno o lo otro tiene lugar *realmente:* R ha actuado y la probabilidad de cada alternativa es de 50% (puesto que es un espejo semirreflectante). Ahora adoptemos el punto de vista de un físico en el *exterior* de la habitación. Podemos suponer que el vector de estado *inicial* de su contenido le es "conocido" antes de que la habitación sea cerrada. (No quiero decir que pudiera saberlo en la práctica, sino que no hay nada en la mecánica cuántica que diga que no pudiera saberlo *en principio.)* Según el observador exterior ninguna "medición" ha tenido lugar

realmente, de modo que la evolución completa del vector de estado debería haber continuado según U. El fotón es emitido por su fuente en su estado predeterminado —ambos observadores coincidirán en esto— y su función de onda se desdobla en dos haces, con una amplitud de, pongamos por ejemplo, $1/\sqrt{2}$) de que el fotón esté en cada uno de ellos (de modo que el cuadrado del módulo diera realmente una probabilidad de 1/2). Puesto que todo el contenido está siendo tratado como un solo sistema cuántico por el observador exterior, la superposición lineal de opciones debe mantenerse hasta la escala del gato. Hay una amplitud $1/\sqrt{2}$ de que la fotocélula registre y una amplitud $1/\sqrt{2}$ de que no lo haga. *Ambas* opciones deben estar presentes en el estado, con el mismo peso como parte de una superposición lineal cuántica. Según el observador exterior el gato está en una superposición lineal de estar muerto y estar vivo. ¿Creemos realmente que sería así? El propio Schrödinger dejó claro

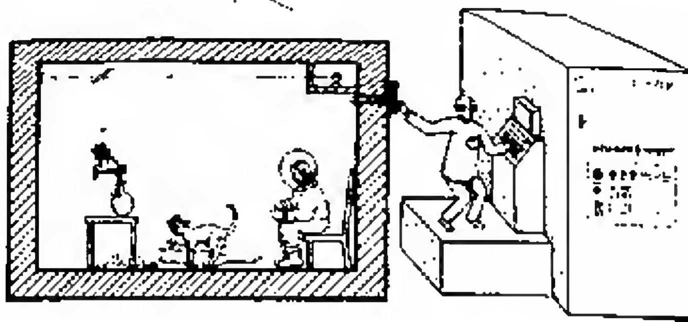


FIGURA VI.33. *El gato de Schrödinger con aditamentos.*

que él no lo creía. Argumentaba, de hecho, que la regla **U** de la mecánica cuántica no sería aplicable a algo tan grande o complicado como un gato. Algo debe haber ido mal con la ecuación de Schrödinger en el camino. Por supuesto, Schrödinger tenía derecho a razonar de esta forma sobre su propia ecuación, pero no es una prerrogativa que se nos conceda a los demás. Muchos físicos (probablemente la mayoría) mantendrían que, por el contrario, hay ahora tanta evidencia experimental a favor de **U** —y absolutamente ninguna en contra— que no tenemos ningún derecho a abandonar ese tipo de evolución, incluso en la escala de un gato. Si se acepta esto, entonces parece que estamos llevados a una visión muy *subjetiva* de la realidad física. Para el observador exterior el gato está realmente en una combinación lineal de estar vivo y muerto, y sólo cuando finalmente se abre la habitación colapsará el vector de estado del gato en uno u otro. Por el contrario, para un observador (adecuadamente protegido) dentro de la habitación, el vector de estado del gato habría colapsado mucho antes, y la combinación lineal del observador exterior

$$|\psi\rangle = 1/\sqrt{2}\,\{|\text{muerto}\rangle + |\text{vivo}\rangle\}$$

no tiene importancia. Parece que después de todo el vector de estado está "todo en la mente".

Pero ¿podemos adoptar realmente esta visión subjetiva del vector de estado? Supongamos que el observador exterior hiciera algo mucho más sofisticado que simplemente "mirar" dentro del contenedor. Supongamos que, a partir de su conocimiento del estado inicial del interior de la habitación, utiliza primero algún medio de cálculo importante disponible para él para *calcular,* mediante la ecuación de Schrödinger, cuál debe ser el estado en el interior del contenedor, obteniendo la respuesta ("correcta") $|\psi\rangle$ (donde $|\psi\rangle$ incluye la anterior superposición lineal de un gato muerto y un gato vivo). Supongamos que entonces él realiza ese experimento *concreto* sobre los contenidos que distingue el propio estado $|\psi\rangle$ de cualquier cosa ortogonal a $|\psi\rangle$. (Como

se ha descrito antes, según las reglas de la mecánica cuántica él puede *en principio* realizar dicho experimento, aunque fuera terriblemente difícil en la práctica.) Los dos resultados *"sí,* está en el estado |ψ⟩" y "no, es ortogonal a |ψ⟩" tendrían probabilidades respectivas del 100 y el 0%. En particular existe una probabilidad nula para el estado |χ⟩ = |muerto⟩−|vivo⟩, que *es* ortogonal a |ψ⟩. La imposibilidad de |ψ⟩ como un resultado del experimento puede surgir sólo a causa de que *ambas* opciones |muerto⟩ y |vivo⟩ *coexisten e* interfieren entre sí.

Lo mismo sería cierto si ajustásemos ligeramente las longitudes de los caminos del fotón (o la proporción reflectante del espejo) de modo que, en lugar del estado |muerto⟩+|vivo⟩ tuviéramos alguna otra combinación, pongamos por caso |muerto⟩−i|vivo⟩, etc. Todas estas diferentes combinaciones tienen consecuencias experimentales distintas en principio. De modo que ya no se trata "simplemente" de cierto tipo de coexistencia entre muerte y vida que pudiera estar afectando a nuestro pobre gato. Todas las diferentes combinaciones *complejas* están permitidas, y todas ellas son, en principio, distinguibles una de otra. Sin embargo, para el observador en el interior de la habitación todas estas combinaciones parecen irrelevantes. O el gato *está* vivo o *está* muerto. ¿Cómo se puede entender este tipo de discrepancia? Señalaré en seguida algunos puntos de vista diferentes que se han manifestado sobre estas (y otras relacionadas) cuestiones, aunque sin duda no haré justicia a todas.

DIVERSAS ACTITUDES HACIA LA TEORÍA CUÁNTICA EXISTENTE

En primer lugar, existen dificultades obvias para realizar un experimento como el que distingue el estado |ψ⟩ de cualquier otro ortogonal a |ψ⟩. No hay duda de que un experimento semejante resulta *en la práctica* imposible para el observador externo. En particular, éste necesitaría conocer el vector de estado exacto de todo el contenido (incluyendo el observador interno) antes de poder siquiera empezar a computar cuál sería realmente |ψ⟩ en un tiempo posterior. Sin embargo, lo que nosotros exigimos es que este experimento sea imposible *en principio* —no simplemente en la práctica— puesto que de otro modo no tendríamos derecho a eliminar uno de los estados "|vivo⟩" o "|muerto⟩" de la realidad física. El problema es que la teoría cuántica, tal como está, no dice cómo trazar una línea clara entre medidas que son "posibles" y las que son imposibles". Tal vez *debería* existir esta distinción tajante. Pero la teoría no lo permite. Para introducir una distinción semejante habría que *cambiar* la teoría cuántica. En segundo lugar, existe el punto de vista no poco frecuente de que las dificultades desaparecerían si pudiéramos tener en cuenta de una manera adecuada el *entorno.* En realidad sería una imposibilidad práctica aislar *efectivamente* por completo el contenido del mundo externo. En cuanto el medio externo se mezcla con el estado en el interior de la habitación, el observador externo no puede considerar los contenidos como dados simplemente por un solo vector de estado. Incluso su *propio* estado se correlaciona con el entorno de una forma complicada. Además, habrá un enorme número de partículas inextricablemente entremezcladas y los efectos de las diferentes combinaciones lineales posibles se extenderán más y más por el universo sobre grandes números de grados de libertad. No hay modo *práctico* (por ejemplo, mediante la observación de efectos de interferencia adecuados) de distinguir estas superposiciones lineales complejas de las simples opciones con pesos probabilistas. Esto ni siquiera tiene que ver con el aislamiento de los contenidos respecto al exterior. El propio gato incluye un gran número de partículas. En consecuencia, la combinación lineal compleja de un gato muerto y uno vivo debe ser tratada *como si* fuera simplemente una mezcla de probabilidades. Sin embargo, yo no encuentro esto nada

satisfactorio. Como sucede con la concepción anterior podemos preguntar en qué etapa se juzga oficialmente que es "imposible" obtener efectos de interferencia, de modo que ahora pueda considerarse que los cuadrados de los módulos de las amplitudes en la superposición compleja proporcionan una probabilidad ponderada de "muerto" y "vivo". Incluso si la "realidad" del mundo se transforma, en cierto sentido, *realmente* en un peso probabilístico en forma de número real, ¿cómo se resuelve esto en una sola alternativa o la otra? Yo no veo siquiera que la *realidad* pueda transformarse sobre la sola base de la evolución **U**, de una *superposición* lineal compleja (o real) de dos opciones en *una o la otra* de estas opciones. Parecemos llevados de nuevo a una visión subjetiva del mundo.

Hay quienes adoptan la postura de que los sistemas complicados no deberían describirse en realidad mediante "estados" sino mediante una generalización conocida como *matrices densidad* (Von Neumann, 1955). Éstas incluyen tanto probabilidades clásicas como amplitudes cuánticas. En efecto, muchos estados cuánticos diferentes se consideran en conjunto para representar la realidad. Las matrices densidad son útiles, pero no resuelven por ellas mismas los profundos problemas de la medida cuántica.

Podríamos tratar de adoptar la postura de que la evolución real es la **U** determinista, pero que las probabilidades surgen de las incertidumbres envueltas en saber cuál *es* realmente el estado cuántico del sistema combinado. Esto sería adoptar una visión muy "clásica" sobre el origen de las probabilidades: la de que todas surgen de las incertidumbres en el estado inicial. Podríamos imaginar qué diferencias pequeñísimas en el estado inicial podrían dar lugar a diferencias enormes en la evolución, como el "caos" que puede ocurrir en los sistemas clásicos *(v.g.* la predicción del tiempo meteorológico; *cfr.* capítulo V). Sin embargo, tales efectos de "caos" sencillamente no pueden ocurrir con **U** por sí sólo, puesto que es *lineal:* las superposiciones lineales no deseadas persisten para siempre con **U**. Para resolver una superposición tal en una alternativa o la otra, se necesita algo no-lineal, de modo que **U** solo no basta.

Para otro punto de vista, debemos tomar nota del hecho de que la única discrepancia completamente clara con la observación, en el experimento del gato de Schrödinger, parece surgir debido a que hay *observadores conscientes,* uno (o dos) en el interior y otro en el exterior de la habitación. Quizá las leyes de la superposición lineal cuántica compleja *no* se aplican a la conciencia. Un modelo matemático aproximado para este punto de vista fue propuesto por Eugene P. Wigner (1961). Él sugirió que la linealidad de la ecuación de Schrödinger podría fallar para entes conscientes (o simplemente "vivientes"), y debería ser reemplazada por algún procedimiento no lineal de acuerdo con el cual se resolvería una u otra alternativa. Pudiera parecer al lector que, puesto que estoy buscando algún tipo de papel para los fenómenos cuánticos en nuestro pensamiento consciente —como de verdad lo estoy haciendo— debería considerar este punto de vista como una atractiva posibilidad. Sin embargo, no me satisface en absoluto. Parece que lleva a una visión muy sesgada y perturbadora de la *realidad* del mundo. Los rincones del universo en donde reside la conciencia podrían ser más bien pocos y muy apartados. Desde esta perspectiva, solo en aquellos rincones podrían resolverse las superposiciones lineales cuánticas complejas en opciones reales. Puede ser que otros rincones semejantes tuvieran, para *nosotros,* la misma apariencia que el resto del universo, puesto que donde quiera que nosotros mismos *miráramos* (u observáramos de algún modo) haríamos, por el mismo acto de observación consciente, que se "resolviese en opciones", *ya lo hubiese hecho antes* o *no.* Sea como fuere, este fuerte sesgo proporcionaría una imagen muy perturbadora de la *realidad* del mundo, y yo, por lo menos, lo aceptaría sólo si me viera obligado a ello.

Hay un punto de vista, relacionado en parte con el anterior, llamado el universo participatorio (sugerido por John A. Wheeler, 1983), que lleva el papel de la conciencia a un (diferente) extremo. Notamos, por ejemplo, que la evolución de la vida consciente en nuestro planeta se debe a mutaciones apropiadas que han tenido lugar en distintos momentos Estas, presumiblemente, son sucesos cuánticos, de modo que sólo existirán en forma linealmente superpuesta hasta que finalmente conduzcan a la evolución de un ser consciente, cuya misma existencia depende de todas las mutaciones correctas que han tenido lugar "realmente" Es nuestra propia presencia la que, en esta concepción, conjura a nuestro pasado a la existencia. La circularidad y paradoja que implica esta idea tiene atractivo para algunos, pero, por mi parte, la encuentro bastante preocupante —y, en realidad, escasamente creíble.

Otro punto de vista, también lógico a su manera pero que proporciona una imagen no menos extraña, es el de los *muchos universos,* publicado por primera vez por Hugh Everett III (1957). Según la interpretación de los muchos universos, **R** no tiene lugar nunca en absoluto. La evolución completa del vector de estado —que se considera de modo realista— está gobernada siempre por el procedimiento determinista **U**. Esto implica que el pobre gato de Schrödinger, junto con el observador protegido en el interior del contenedor, debe existir en alguna combinación lineal compleja con el gato en alguna superposición de vida y muerte. Sin embargo, el estado muerte está correlacionado con un estado de la conciencia del observador interno, y el estado vivo, con otro (y presumiblemente, en parte, con la conciencia del gato y, en última instancia, también con el observador externo cuando el contenido le sea revelado). La conciencia de cada observador se "desdobla", de modo que ahora existe por duplicado, y cada uno de sus ejemplares tiene una experiencia diferente (es decir, uno que ve un gato vivo y otro que ve un gato muerto). De hecho, no sólo un observador sino todo el universo en el que habita se desdobla en dos (o más) en cada medición que hace del mundo. Este desdoblamiento ocurre una y otra vez —no simplemente a causa de "medidas" hechas por observadores, sino a causa de la amplificación macroscópica de estados cuánticos en general— de modo que estas "ramas" de universo proliferan incontroladamente. En realidad, todas las posibilidades opcionales coexistirán en una vasta superposición. Este no es precisamente el más económico de todos los puntos de vista, pero mis propias objeciones a él no derivan de su falta de economía. En particular, no veo por qué un ser consciente necesita ser consciente de sólo "una" de las opciones en una superposición lineal. ¿Qué conciencia es esa que exige que no podamos ser "conscientes" de esa seductora combinación lineal de un gato muerto y un gato vivo? Creo que sería necesaria una teoría de la conciencia antes de que la idea de los muchos universos pueda ser confrontada con lo que realmente observamos. No veo qué relación existe entre el vector de estado "verdadero" (objetivo) del universo y el que se supone que observamos "realmente". Se ha expuesto que la "ilusión" de **R** puede, en cierto sentido, deducirse efectivamente en esta imagen pero no creo que estas afirmaciones se sostengan. Cuando menos se necesitan más ingredientes para que el esquema funcione. Creo que la idea de los muchos universos introduce una multitud de problemas propios sin resolver realmente los *auténticos* enigmas de la medida cuántica. (Compárese con De Witt y Graham, 1973.)

¿DÓNDE NOS DEJA TODO ESTO?

Estos enigmas persisten, de una u otra forma, en *cualquier* interpretación de la mecánica cuántica tal como hoy existe. Repasemos brevemente lo que la teoría cuántica estándar nos ha dicho

realmente acerca de cómo debemos describir el mundo, especialmente en relación con estos intrigantes temas; y luego preguntemos: ¿a dónde vamos desde aquí?

Recordemos, antes de nada, que las descripciones de la teoría cuántica parecen aplicarse acertadamente (¿útilmente?) sólo al llamado *nivel cuántico* de moléculas, átomos o partículas subatómicas, pero también a dimensiones mayores siempre que las diferencias de energía entre posibilidades opcionales permanezcan muy pequeñas. En el nivel cuántico debemos tratar tales "opciones" como cosas que pueden *coexistir* en alguna especie de superposición con pesos estadísticos complejos. Los números complejos que se utilizan como pesos se llaman *amplitudes de probabilidad.* Cada diferente totalidad de opciones con pesos complejos define un diferente *estado cuántico,* y cualquier sistema cuántico debe describirse mediante uno de estos estados cuánticos. Con frecuencia, como sucedía muy claramente con el ejemplo del spin, no hay nada qué decir sobre cuáles son las opciones "reales" que componen un estado cuántico y cuáles son sólo "combinaciones" de opciones. En cualquier caso, mientras el *sistema permanezca* en el nivel cuántico, el estado cuántico evoluciona de una forma completamente *determinista.* Esta evolución determinista es el proceso **U**, gobernado por la importante *ecuación de Schrödinger,*

Cuando los efectos de diferentes opciones cuánticas se amplifican hasta el *nivel clásico,* de modo que las diferencias entre las opciones son bastante grandes para que podamos percibirlas directamente, entonces esas superposiciones con pesos complejos ya no parecen persistir más. En su lugar, deben formarse los cuadrados de los módulos de las amplitudes complejas (es decir, tomar los cuadrados de sus distancias al origen en el plano complejo), y estos números *reales* juegan ahora un nuevo papel *como probabilidades* reales para las opciones en cuestión. Sólo una de las opciones sobrevive en la realidad de la experiencia física siguiendo el proceso **R** (llamado reducción del vector de estado o colapso de la función de onda y que es completamente diferente de **U**). Es aquí y solo aquí, donde hace su entrada el no determinismo de la teoría cuántica

Puede defenderse con fuerza que el estado cuántico proporciona una imagen *objetiva.* Pero puede ser una imagen complicada y algo paradójica. Cuando varias partículas están involucradas, los estados cuánticos pueden (y normalmente lo "hacen") hacerse muy complicados. En tal caso, las partículas individuales no tienen "estados" por sí mismas sino que existen solamente en complicados "entramados" con otras partículas, conocidos como *correlaciones.* Cuando una partícula en una región es "observada", en el sentido de que desencadena algún efecto que se amplifica hasta el nivel clásico, entonces debe acudirse a R, pero en apariencia esto afecta *simultáneamente* a todas las demás partículas con la que esta partícula concreta está correlacionada. Los experimentos del tipo Einstein-Podolsky-Rosen (EPR) (como el de Aspect, en el que una fuente cuántica emite un par de fotones en direcciones opuestas y luego se miden independientemente sus polarizaciones cuando están a muchos metros de distancia) dan una sustancia observacional evidente a este enigmático aunque esencial hecho de la física cuántica: ésta es *no local* (de modo que los fotones en el experimento de Aspect no pueden tratarse como entidades independientes). Si se considera que R actúa de una manera objetiva (y eso parecería estar implicado en la objetividad del estado cuántico) entonces el espíritu de la relatividad especial es violado en consecuencia. No parece existir *ninguna descripción espacio-temporal objetivamente real* del vector de estado (que se reduce) que sea consistente con los requisitos de la relatividad. Sin embargo, los efectos *observacionales* de la teoría cuántica no violan la relatividad.

La teoría cuántica guarda silencio sobre *cuándo* y *por qué* R debería tener lugar (¿o aparentarlo?). Además, no explica adecuadamente, por sí sola, porque el mundo del nivel clásico "parece" clásico. "La mayoría" de los estados cuánticos no se parece en absoluto a los clásicos.

¿Dónde nos deja todo esto? Creo que debemos considerar seriamente la posibilidad de que la mecánica cuántica sea sencillamente *errónea* cuando se aplica a cuerpos macroscópicos o, más bien que las leyes U y R sólo suministran excelentes aproximaciones a alguna teoría más completa aunque todavía desconocida. Es la *combinación* de estas dos leyes juntas, y no U por sí sola, la que ha proporcionado todo el maravilloso acuerdo con la observación de que goza la teoría actual. Si se extendiera la linealidad de U al mundo macroscópico tendríamos que aceptar la realidad física de combinaciones lineales complejas de diferentes posiciones (o diferentes spines, etc.) de bolas de cricket y similares. El simple sentido común nos dice que el mundo no se comporta realmente de este modo. Las bolas de cricket se aproximan muy bien mediante las descripciones de la física *clásica*. Tienen posiciones razonablemente bien definidas, y no se las ve en dos lugares a la vez, como les permitirían estar las leyes de la mecánica cuántica. Si hay que reemplazar los procedimientos **U** y **R** por una ley más amplia, entonces, a diferencia de la ecuación de Schrödinger, esta nueva ley tendría que ser de carácter *no* lineal (ya que el propio **R** actúa de forma no lineal). Algunos presentan objeciones a esto, apuntando muy correctamente que buena parte de la elegancia matemática de la teoría cuántica estándar es resultado de su linealidad. Sin embargo, me sentiría sorprendido si la teoría cuántica no fuera a tener algún cambio fundamental en el futuro, hacia algo para lo que esta linealidad fuera sólo una aproximación. Ciertamente hay precedentes para este tipo de cambio. La poderosa y elegante teoría de Newton de la gravitación universal debía mucho al hecho de que las fuerzas de la teoría se suman de una manera *lineal*. Pero con la teoría de la relatividad general de Einstein se vio que esta linealidad era sólo una aproximación (si bien excelente), y la elegancia de la teoría de Einstein supera incluso a la de Newton.

No me he andado con rodeos sobre el hecho de que creo que la resolución de los enigmas de la mecánica cuántica debe estar en el descubrimiento de una teoría mejorada. Aunque quizá no sea esta la opinión convencional tampoco es completamente no convencional. (Muchos de quienes dieron origen a la teoría cuántica eran también de este parecer. He citado las opiniones de Einstein. Schrödinger [1935], de Broglie [1956] y Dirac [1939] también consideraban provisional la teoría.) Pero incluso si se piensa que la teoría debe ser modificada de alguna manera, las restricciones sobre *cómo* hacerlo son enormes. Quizá algún tipo de punto de vista de "variables ocultas" resultaría aceptable finalmente. Pero la no localidad que muestran los experimentos de tipo EPR desafían seriamente cualquier descripción "realista" del mundo que pueda ocurrir cómodamente en un espacio-tiempo ordinario —un espacio-tiempo del tipo particular que nos ha sido dado para coincidir con los principios de la relatividad—, así que creo que se necesita un cambio mucho más radical. Además, nunca se ha encontrado ninguna discrepancia de ningún tipo entre la teoría cuántica y los experimentos, a menos, por supuesto, que consideremos como evidencia en contra la ausencia de bolas de cricket superpuestas linealmente. En mi opinión, la no existencia de bolas de cricket superpuestas linealmente *es* ciertamente evidencia en contra. Pero esto, en sí mismo, no es de gran ayuda. Sabemos que en el nivel submicroscópico de las cosas las leyes cuánticas son válidas; pero en el nivel de las bolas de cricket es la física clásica la que vale. Sostendré que en algún lugar intermedio necesitamos comprender la nueva ley para ver como el mundo cuántico enlaza con el clásico. Creo también

que necesitamos esta nueva ley si queremos conocer alguna vez las mentes. Por todo esto pienso que debemos buscar nuevas claves.

En mis descripciones de la teoría cuántica en este capítulo he sido totalmente convencional, aunque con un énfasis quizá más geométrico y "realista" de lo que es usual. En el próximo capítulo trataremos de buscar algunas claves necesarias —claves que creo que nos pueden dar algunas ideas sobre una mecánica cuántica mejorada—. Nuestro viaje se iniciará cerca de casa, pero nos veremos obligados a viajar muy lejos. Resulta que necesitaremos explorar dominios del espacio muy diferentes, y retroceder incluso al propio comienzo del tiempo.

VII LA COSMOLOGÍA Y LA FLECHA DEL TIEMPO

EL FLUJO DEL TIEMPO

LA SENSACIÓN DEL PASO DEL TIEMPO es central para nuestros sentimiento de conciencia. *Parece* que nos estamos moviendo siempre hacia adelante, desde un pasado definido hacia un futuro incierto. Sentimos que el pasado se ha ido y ya no hay nada que hacer con él. No se puede cambiar y, en cierto sentido, todavía está "ahí fuera". Nuestro conocimiento presente de él puede proceder de nuestros registros, de nuestra memoria y de nuestras deducciones de éstos, pero no tenemos tendencia a dudar de la *realidad* del pasado. El pasado era una cosa y sólo puede *ser* (ahora) una cosa. Lo que sucedió, sucedió y ni nosotros, ni nadie más, podemos hacer nada por cambiarlo! Por el contrario, el futuro parece aún indeterminado. Podría resultar ser una cosa o podría ser otra. Quizá esta "elección" está completamente determinada por las leyes físicas, o quizá en parte por nuestras propias decisiones (o por Dios); pero *parece* que esta "elección" está aún por hacerse. Da la impresión de que sólo hay simples *potencialidades* para cualquier cosa que la "realidad" del futuro pueda decidirse a ser. A medida que percibimos conscientemente que el tiempo pasa, la parte más inmediata de este vasto y aparentemente indeterminado futuro se realiza como actualidad y, de este modo, hace su entrada en el pasado fijo. A veces podemos tener la sensación de que *nosotros* hemos sido incluso personalmente "responsables" al influir de algún modo en esta elección del futuro potencial concreto que se realiza efectivamente y se hace permanente en la realidad del pasado. Más frecuentemente nos sentimos como espectadores inútiles —quizá agradecidos por este alivio de responsabilidad— de cómo, inexorablemente, la frontera del pasado determinado se mueve hacia el futuro incierto.

Pese a ello, la física, tal como la conocemos, nos cuenta una historia diferente. Todas las ecuaciones fructíferas de la física son simétricas respecto al tiempo. Se pueden utilizar tanto en una dirección del tiempo como en la otra. El futuro y el pasado parecen estar físicamente en pie de igualdad. Las leyes de Newton, las ecuaciones de Hamilton, las ecuaciones de Maxwell, la relatividad general de Einstein, la ecuación de Dirac, la ecuación de Schrödinger..., todas permanecen inalteradas si invertimos la dirección del tiempo (lo que equivale a reemplazar la coordenada *t,* que representa el tiempo, por *-t).* Toda la mecánica clásica

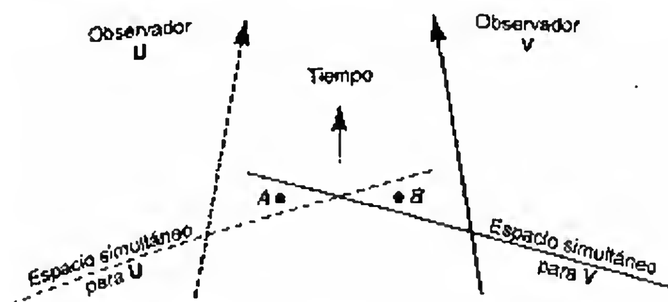


FIGURA VII.1 *¿Puede el tiempo "fluir" realmente? Para un observador* ***U****,* B *puede estar en el pasado "fijo" mientras que* A *está todavía en el futuro "incierto". El observador* ***V*** *mantendrá la opinión contraria.*

junto con la parte "**U**" de la mecánica cuántica, es completamente reversible en el tiempo. Hay una cuestión abierta sobre si la parte "**R**" de la mecánica cuántica es realmente reversible o no. Esta cuestión será capital para los argumentos que presentaré en el próximo capítulo. Por el momento, dejemos de lado este tema y refirámonos a lo que puede considerarse el "saber

convencional" sobre la materia, según el cual, a pesar de las apariencias iniciales, la operación de R debe ser aceptada también como tiempo-simétrica *(cfr.* Aharonov, Bergmann y Lebowitz, 1964). Si aceptamos esto, parece que tendremos que buscar en otra parte si queremos encontrar el lugar donde, según nuestras leyes físicas, debe residir la diferencia entre pasado y futuro.

Antes de abordar este punto, consideremos otra discrepancia intrigante entre nuestra percepción del tiempo y lo que la teoría física moderna nos dice que debemos creer. Según la relatividad, no existe en absoluto algo como el "ahora". Lo más cercano que tenemos de tal concepto es el espacio simultáneo" de un observador en el espacio-tiempo, como se representa en la fig. V.21, pero éste depende del *movimiento* del observador. El "ahora" de un observador no coincidiría con el de otro.[1] Con respecto a dos sucesos espacio-temporales *A* y *B,* un observador **U** Podría considerar que *B* pertenece al pasado fijo y *A* al futuro incierto, mientras que para un segundo observador **V**, ¡sería *A* el que perteneciera al pasado fijo y *B* al futuro incierto! (véase fig. VII 1). No tiene sentido afirmar que uno cualquiera de los sucesos *A* y *B* permanece incierto en tanto que el otro está definido.

Recordemos lo expuesto y la fig. V.22. Dos personas se cruzan en la calle: según una de ellas, una flota espacial de Andrómeda ha iniciado ya su viaje, mientras que, según la otra, todavía no se ha tomado la decisión de realizar o no dicho viaje. ¿Cómo puede haber todavía alguna incertidumbre sobre el resultado de dicha decisión? Si para *cualquiera* de las dos personas la decisión ya ha sido tomada, entonces es seguro que *no puede haber* ninguna incertidumbre. El lanzamiento de la flota espacial es algo inevitable. En realidad ninguna de las dos personas puede *saber* todavía del lanzamiento de la flota espacial. Ellas sólo lo pueden saber más tarde, cuando las observaciones telescópicas desde la Tierra revelen que la flota está realmente en camino. Entonces ellas podrían recordar su encuentro,[2] y llegar a la conclusión de que en *ese* momento, según una de ellas, la decisión estaba en el futuro incierto, mientras que según la otra, estaba en el pasado cierto. ¿Había *entonces* alguna incertidumbre sobre dicho futuro? ¿O estaba "fijo" ya el futuro de *ambas* personas? Empieza a parecer que basta con que algo esté definido para que todo el espacio-tiempo deba estar definido. No puede haber futuro "incierto". La *totalidad* del espacio-tiempo debe estar fijada, sin ninguna perspectiva para la incertidumbre. De hecho, ésta parece haber sido la propia conclusión de Einstein *(cfr.* País, 1982, p. 444). Además, no hay flujo del tiempo en absoluto. Sólo tenemos "espacio-tiempo", y ninguna perspectiva para un futuro cuyo dominio está siendo invadido inexorablemente por un pasado determinado! (El lector puede estar preguntándose cuál es el papel de las "incertidumbres" de la mecánica cuántica en todo esto. Volveré más adelante en el próximo capítulo a las cuestiones que plantea la mecánica cuántica. Por el momento será mejor que pensemos en términos de una imagen puramente clásica.) Tengo la impresión de que hay serias discrepancias entre lo que sentimos conscientemente, con relación al flujo del tiempo, y lo que nuestras teorías (maravillosamente precisas) afirman sobre la realidad del mundo físico. Seguramente estas discrepancias nos están diciendo algo profundo acerca de la física que presumiblemente debe subyacer a nuestras percepciones conscientes, suponiendo (como creo) que lo que subyace a estas percepciones sea inteligible mediante algún tipo apropiado de física. Al menos parece evidente que, cualquiera que sea la física que esté actuando, ella debe tener un ingrediente esencialmente tiempo-

---

[1] Algunos "Puristas" de la relatividad prefieren utilizar los conos de luz de los observadores mas que sus espacios simultáneos. Sin embargo, esto no supone la más mínima conclusión.

[2] Se me ocurrió, al revisar este escrito, que para entonces ambas personas ya habrían muerto hace mucho tiempo; serían sus *descendientes lejanos* quienes tendrían que "recordar" el encuentro.

asimétrico, esto es, debe hacer una distinción entre el pasado y el futuro. Si las ecuaciones de la física no parecen hacer distinción entre futuro y pasado —y si incluso la misma idea del "presente" no encaja perfectamente en la relatividad, entonces ¿dónde tenemos que buscar para encontrar leyes físicas mas en consonancia con lo que parecemos percibir del mundo? En realidad, las cosas no son tan discrepantes como parece que he estado dando a entender. Nuestro conocimiento de la física contiene realmente ingredientes importantes *además* de las solas ecuaciones de evolución temporal, y algunos de éstos implican asimetrías temporales. El más importante de éstos es lo que se conoce como la *segunda ley de la termodinámica*. Intentemos sacar alguna idea de lo que esta ley significa.

EL INCREMENTO INEXORABLE DE LA ENTROPÍA

Imaginemos un vaso de agua en equilibrio en el borde de una mesa. Si se le empuja ligeramente es probable que el vaso caiga al suelo, sin duda para hacerse añicos, con el agua salpicada por un área considerable, quizá absorbida en una alfombra o en las ranuras entre las baldosas del suelo. En esto, nuestro vaso de agua no ha hecho más que seguir fielmente las ecuaciones de la física. Las descripciones de Newton bastarán. Cada uno de los átomos del vaso y del agua está obedeciendo individualmente las leyes de Newton (fig. VII.2). Ahora hagamos pasar estas imágenes al revés en la dirección inversa del tiempo. Por la reversibilidad temporal de estas leyes, el agua podría brotar de la alfombra y de las ranuras entre las baldosas, meterse en el vaso que está ocupado en construirse a sí mismo a partir de los numerosos trozos dispersos, para saltar luego todo el conjunto desde el suelo hasta la altura exacta de la mesa y colocarse en reposo equilibrado en el borde de la mesa. Todo esto está de acuerdo con las leyes de Newton, lo mismo que la caída y rotura del vaso.

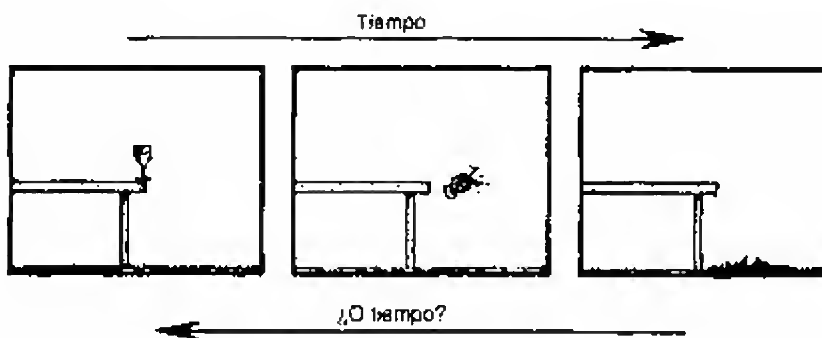


FIGURA VII.2. *Las leyes de la mecánica son reversibles respecto al tiempo; pese a ello nunca se tiene experiencia de una escena con un orden temporal que vaya desde la imagen derecha hacia la izquierda, mientras que la escena con orden temporal de izquierda a derecha sería algo muy común.*

El lector puede estar preguntándose quizá de dónde sale la energía que eleva el vaso desde el suelo a la mesa. *Esto* no es problema. No puede haber problema con la energía puesto que en la situación en la que el vaso *cae* desde la mesa, la energía que adquiere de la caída debe *ir* a otra parte. De hecho, la energía del vaso que cae se transforma en *calor*. Los átomos de los fragmentos del vaso, del agua, de la alfombra y de las baldosas se están moviendo de manera aleatoria justo un poco más rápidos de lo que lo hacían en el momento en que el vaso chocó contra el suelo, es decir, los fragmentos del vaso, agua, alfombra y baldosas estarán apenas un

poco *más calientes* de lo que estaban antes (ignorando posibles pérdidas de calor por evaporación, aunque también eso es reversible en principio). Por la *conservación de la energía,* esta energía calorífica es exactamente igual a la energía perdida por el vaso de agua al caer desde la mesa. Por lo tanto, esta pequeña cantidad de energía calorífica será justo suficiente para elevar el vaso de nuevo hasta la mesa. Es importante darse cuenta de que debe ser incluida la energía calorífica cuando consideramos la conservación de la energía. La ley de la conservación de la energía, cuando se tiene en cuenta la energía calorífica, se llama *primera ley de la termodinámica.* Al ser una deducción de la mecánica newtoniana, la primera ley de la termodinámica es tiempo-simétrica. La primera ley *no* impone ninguna limitación al vaso y al agua que excluya su recomposición, el llenado con agua y el salto milagroso a la mesa.

La razón de que no veamos que suceden estas cosas es que el movimiento "térmico" de los átomos en los fragmentos del vaso, agua, baldosas y alfombra será completamente desordenado, de modo que la mayoría de los átomos se estarán moviendo en las direcciones inadecuadas. Sería necesaria una coordinación absurdamente precisa de sus movimientos para recomponer el vaso, con todas las salpicaduras de agua recuperadas dentro, y alzarlo delicadamente hasta la mesa. Es una certeza efectiva que semejante movimiento coordinado *no* estará presente! Semejante coordinación sólo podría ocurrir mediante el más inverosímil golpe de azar, de un tipo que llamaríamos "mágico" si se diera alguna vez.

Sin embargo, en la otra dirección del tiempo semejante movimiento coordinado es un lugar común. De ningún modo consideramos un golpe de suerte que las partículas se muevan de forma coordinada, siempre que lo hagan *después* de que haya tenido lugar algún cambio a gran escala en el sistema físico (en este caso la ruptura y derramamiento del vaso de agua), y no *antes* de tal cambio. Los movimientos de las partículas deben efectivamente estar fuertemente coordinados después de tal suceso, pues estos movimientos son de tal naturaleza que si tuviéramos que invertir de una manera exacta el movimiento de cada átomo individual, el comportamiento resultante sería exactamente el necesario para recomponer, llenar y levantar el vaso hasta su exacta configuración de partida.

El movimiento fuertemente coordinado es aceptable si se considera como un *efecto* de un cambio a gran escala, y no como la *causa* de éste. Sin embargo, las palabras "causa" y "efecto" encierran de algún modo una petición de principio sobre el problema de la asimetría temporal. En nuestra forma de hablar normal acostumbramos a aplicar estos términos en el sentido de que la causa debe preceder al efecto. Pero, si estamos tratando de comprender la diferencia física entre pasado y futuro, tenemos que tener mucho cuidado de no introducir inconscientemente en la discusión nuestras sensaciones cotidianas sobre pasado y futuro. Debo advertir al lector que es extremadamente difícil evitarlo, pero es imperativo que tratemos de hacerlo. Debemos tratar de utilizar las palabras de tal modo que no prejuzguen el resultado de la diferencia física entre pasado y futuro. En consecuencia, si las circunstancias lo juzgan apropiado tendríamos que permitirnos tomar las causas de las cosas como estando en el futuro y los efectos como estando en el pasado. Las ecuaciones deterministas de la física clásica (o, en su caso, la operación de U en la física cuántica) no tienen preferencia por evolucionar en la dirección del futuro. Pueden utilizarse igualmente para evolucionar hacia el pasado. El futuro determina el pasado exactamente de la misma forma que el pasado determina el futuro. Podemos especificar un estado de un sistema de algún modo arbitrario en el futuro y luego utilizar este estado para calcular cómo hubiera debido ser en el pasado. Si se nos permite ver el pasado como "causa" y el futuro como "efecto", cuando seguimos las ecuaciones del sistema en la dirección normal del

futuro, entonces cuando aplicamos el procedimiento igualmente válido de seguir las ecuaciones en la dirección del pasado debemos considerar aparentemente el futuro como "causa" y el pasado como "efecto".

Sin embargo, hay algo más implícito en nuestro uso de los términos "causa" y "efecto" que no es cuestión realmente de cuál de los sucesos referidos esté en el pasado y cuál en el futuro. Imaginemos un universo hipotético en el que se aplican las mismas ecuaciones clásicas con simetría temporal que en nuestro propio universo, pero en el que el comportamiento de tipo familiar (*v.g.* la rotura y derramamiento de los vasos de agua) coexiste con aconteceres como los inversos de éstos en el tiempo. Supongamos que, a la par con nuestras experiencias más familiares, los vasos de agua a veces se *recomponen* a partir de los pedazos rotos, se llenan misteriosamente a partir de salpicaduras de agua y luego trepan a las mesas; supongamos también que, en ocasiones, los huevos revueltos se separan y desfríen mágicamente, para volver finalmente a sus cáscaras rotas que se recomponen perfectamente y se cierran; que los terrones de azúcar pueden formarse por sí solos a partir del azúcar disuelto en un café azucarado y luego saltan espontáneamente desde la taza hasta la mano de alguien. Si viviéramos en un mundo en el que tales aconteceres fueran un lugar común, seguramente atribuiríamos las "causas" de tales sucesos no a coincidencias azarosas fantásticamente improbables respecto al comportamiento correlacionado de los átomos individuales, sino a algún "efecto teleológico" por el que los objetos autoformantes luchan a veces por conseguir alguna configuración macroscópica deseada. "¡Mira!", diríamos, "está ocurriendo otra vez. ¡Este revoltijo va a recomponerse en otro vaso de agua!" Sin duda aceptaríamos la idea de que los átomos se dirigían a sí mismos de forma tan precisa *debido* a que ésta era la forma de producir el vaso de agua sobre la mesa. El vaso sobre la mesa sería la "causa", y la aparentemente aleatoria colección de átomos en el suelo sería el "efecto", pese al hecho de que el "efecto" ocurre ahora en un tiempo anterior a la "causa". Análogamente, el movimiento minuciosamente organizado de los átomos en el huevo revuelto no es la "causa" del levantamiento hasta la cáscara recompuesta, sino el "efecto" de este acontecer futuro; y el terrón de azúcar no se reúne por sí mismo y se sale de la taza "a causa de que" los átomos se muevan con tan extraordinaria precisión, sino debido al hecho de que alguien —aunque en el futuro— sostendrá más tarde ese terrón de *azúcar* en su mano.

Por supuesto no vemos que tales cosas sucedan en nuestro mundo o, mejor dicho, lo que no vemos es la *coexistencia* de tales cosas con las de nuestro tipo normal. Si *todo* lo que viésemos fueran aconteceres del tipo anómalo recién descrito, entonces no tendríamos problema. Podríamos intercambiar simplemente los términos "pasado" y "futuro", "antes" y "después", etc., en todas nuestras descripciones. El tiempo podría ser considerado avanzando en la dirección inversa de la especificada originalmente, y dicho mundo podría describirse como si fuera exactamente igual que el nuestro. Sin embargo, aquí estoy considerando una posibilidad diferente —pero igual de consistente con las ecuaciones tiempo-simétricas de la física— en la que la ruptura y la auto-recomposición de los vasos de agua pueden *coexistir*. En un mundo semejante no podemos recuperar nuestras descripciones familiares simplemente mediante una inversión de nuestras convenciones sobre la dirección de avance del tiempo. Por supuesto, nuestro mundo no parece ser así, pero ¿por qué no lo es? Para empezar a comprender este hecho he estado pidiéndoles que traten de imaginarse un mundo semejante y preguntarse cómo describiríamos los aconteceres que tienen lugar en él. Les estoy pidiendo que acepten que, en semejante mundo, ciertamente describiríamos las grandes configuraciones macroscópicas —tales como vasos de agua enteros, huevos intactos, o un terrón de azúcar sostenido en una mano—

como si fueran las "causas" , y los detallados, y quizá estrechamente correlacionados, movimientos de los átomos individuales como "efectos", estén o no las "causas" en el futuro o en el pasado de los "efectos"- ¿Por qué, en el mundo que nos ha tocado vivir, son las causas las que *preceden* a los efectos?; o, por poner las cosas de otra forma, ¿por qué los movimientos de partículas exactamente coordinados ocurren sólo *después* de algún cambio a gran escala en el sistema físico y no *antes* de él? Para dar una mejor descripción física de tales cosas necesitaré introducir el concepto de *entropía.* En términos generales, la entropía de un sistema es una medida de su *desorden* manifiesto. (Más adelante seré un poco más preciso.) Así, el cristal roto y el agua desparramada por el suelo están en un estado de mayor entropía que el del vaso entero y lleno de agua en la mesa; el huevo revuelto tiene una entropía más alta que el huevo fresco intacto; el café azucarado tiene una entropía más alta que el terrón de *azúcar* sin disolver en un café amargo. El estado de baja entropía parece "particularmente ordenado", de algún modo manifiesto, y el estado de alta entropía, menos "particularmente ordenado".

Es importante darse cuenta de que cuando nos referimos a la "particularidad" de un estado de baja entropía nos estamos refiriendo en realidad a una particularidad *manifiesta.* De hecho, en un sentido más sutil, el estado de mayor entropía, en estas situaciones, *está* tan "particularmente ordenado" como el estado de menor entropía, debido a la muy precisa coordinación de movimientos de las partículas individuales. Por • ejemplo, los movimientos aparentemente aleatorios de las moléculas de agua que se han escurrido entre las baldosas después de que el vaso se ha roto son realmente muy especiales: los movimientos son tan precisos que si todos ellos fueran *invertidos* exactamente se recobraría el estado original de baja entropía en el que el vaso está sobre la mesa entero y lleno de agua. (Esto debe ser así puesto que la inversión de todos estos movimientos correspondería simplemente a invertir la dirección del tiempo, según lo cual el vaso se recompondría por sí mismo y saltaría a la mesa.) Pero semejante movimiento coordinado de todas las moléculas de agua no es el tipo de "particularidad" que llamamos baja entropía. La entropía se refiere al desorden *manifiesto.* El orden que está presente en la coordinación exacta de movimientos de partículas no es orden manifiesto, así que no cuenta para disminuir la entropía de un sistema. Por lo tanto, el orden en las moléculas del agua desparramada no cuenta en este sentido y la entropía es alta. Sin embargo, el orden *manifiesto* en el vaso de agua *compuesto* da un bajo valor de la entropía. Éste se refiere al hecho de que relativamente pocas configuraciones posibles diferentes de movimientos de partículas son compatibles con la configuración manifiesta de un vaso de agua entero y lleno; mientras que existen muchos más movimientos que son compatibles con la configuración manifiesta del agua ligeramente calentada que fluye entre las ranuras de las baldosas. La *segunda ley de la termodinámica* afirma que *la entropía de un sistema aislado aumenta con el tiempo (o en el caso de un sistema reversible, permanece constante).* Está bien que no contemos los movimientos coordinados de partículas como baja entropía pues si lo hiciéramos, la "entropía" de un sistema, según esa definición, siempre permanecería constante. El concepto de entropía debe referirse sólo al desorden que es realmente manifiesto. Para un sistema aislado del resto del Universo, su entropía total crece, de modo que si se parte de algún estado con algún tipo de organización manifiesta esta organización será erosionada en el curso del tiempo, y estas especiales características manifiestas se convertirán en "inútiles" movimientos coordinados de partículas. Podría parecer, tal vez, que la segunda ley es una especie de último recurso, que afirma que hay un principio físico universal e inexorable que nos dice que la organización se rompe continuamente por necesidad. Veremos más adelante que esta conclusión pesimista no es completamente apropiada.

¿QUÉ ES LA ENTROPÍA?

Pero ¿qué *es* exactamente la entropía de un sistema físico? Hemos visto que es algún tipo de medida del desorden manifiesto, pero podría parecer, dada mi utilización de términos tan imprecisos como "desorden" y "manifiesto", que el concepto de entropía no fuera realmente una cantidad científicamente muy clara. Existe también otro aspecto de la segunda ley que parece señalar un elemento de imprecisión en el concepto de entropía: es sólo en los llamados sistemas *irreversibles* en donde la entropía crece en lugar de permanecer constante. ¿Qué significa "irreversible"? Si tenemos en cuenta los movimientos detallados de todas las partículas, entonces ¡*todos* los sistemas son reversibles! *En la práctica* diríamos que el vaso que cae de la mesa y se rompe, o el revolver del huevo, o la disolución del azúcar en el café son todos irreversibles; mientras que el rebotar de un pequeño número de partículas unas en otras sería considerado reversible, como lo serían diversas situaciones cuidadosamente controladas en las que la energía no se transforma en calor. Básicamente, el término "irreversible" se refiere simplemente al hecho de que no ha sido posible seguir, ni controlar, todos los detalles importantes de los movimientos de las partículas individuales del sistema. Estos movimientos incontrolados son designados como "calor". Así, la irreversibilidad parece ser simplemente una cuestión "práctica". No podemos *en la práctica* desrevolver un huevo, aunque es un procedimiento perfectamente admitido según las leyes de la mecánica. ¿Depende nuestro concepto de entropía de lo que es prácticamente factible y lo que no lo es? Recordemos, del capítulo V, que se *puede* dar una definición matemática precisa del concepto físico de *energía,* tanto como de los de momento o momento angular, en términos de posiciones, velocidades, masas y fuerzas sobre las partículas. Pero ¿cómo podemos esperar que también se pueda hacer esto para el concepto de "desorden manifiesto" que es necesario para hacer matemáticamente preciso el concepto de entropía? Ciertamente, lo que es "manifiesto" para un observador puede no serlo para otro. ¿Dependería de la precisión con que cada observador sea capaz de medir el sistema? Un observador con mejores instrumentos de medida podría obtener una información más detallada acerca de los constituyentes microscópicos de un sistema, que la que pudiera obtener otro. Para un observador podría ser manifiesta una parte mayor del "orden oculto" y, consecuentemente, concebiría una entropía más baja que el otro. Parece también que los juicios estéticos de los diversos observadores estarían involucrados en lo que ellos consideraran que es "orden" más que "desorden". Podríamos pensar incluso en algún artista que opinara que la colección de fragmentos del vaso roto está ordenada de forma mucho más bella de lo que estaba el vaso horriblemente feo que estuvo una vez en el borde de la mesa. ¿Habría sido *reducida* la entropía realmente en el juicio de semejante observador, artísticamente sensible?

A la vista de estos problemas de subjetividad, resulta sorprendente que el concepto de entropía tenga utilidad en descripciones científicamente precisas. Pero ciertamente la tiene. La razón de esa utilidad es que los cambios de orden a desorden en un sistema, en términos de posiciones y velocidades detalladas de las partículas, son enormes y (en casi todas las circunstancias) desbordarán todas las diferencias razonables entre puntos de vista sobre lo que es o no es un "orden manifiesto" a escala macroscópica. En particular, el juicio del artista o el del científico respecto a si es el vaso intacto o el roto el que es un arreglo de mayor orden, no tiene apenas consecuencias respecto a su medida de la entropía. La mayor contribución a la entropía procede del movimiento aleatorio de las partículas que suponen el pequeñísimo incremento de temperatura y la dispersión del agua cuando vaso y agua golpean el suelo.

Para mayor precisión del concepto de entropía, volvamos a la idea de *espacio de fases* que fue introducida en el capítulo V. Recordemos que el espacio de fases de un sistema es un espacio, normalmente de un enorme número de dimensiones, cada uno de cuyos puntos representa un estado físico completo en sus más mínimos detalles. Un *simple* punto en el espacio *de* fases proporciona todas las coordenadas de posición y momento de todas las partículas individuales que constituyen el sistema físico en cuestión. Lo que necesitamos para el concepto de entropía es una manera de agrupar todos los estados que parezcan idénticos desde el punto de vista de sus propiedades *manifiestas* (esto es, macroscópicas). Necesitamos dividir nuestro espacio de fases en un número de compartimentos (véase fig. VII.3), de manera que todos los puntos de un compartimento concreto representan sistemas físicos que —aunque diferentes en los detalles menudos de las configuraciones y movimientos de sus partículas— se consideran idénticos con respecto a sus características macroscópicamente observables. Desde el punto de vista de lo que *es* manifiesto, todos los puntos de un compartimento representan el *mismo* sistema físico. Esta división del espacio de fases en compartimentos se conoce como división de *grano-grueso* del espacio de fases.

Ahora resultará que algunos de estos compartimentos serán inmensamente mayores que otros. Por ejemplo, consideremos el espacio de fases de un gas en una caja. La mayor parte de éste corresponderá a estados en los que el gas está uniformemente distribuido dentro de la caja, con las partículas moviéndose de una forma característica que proporcione una temperatura y presión uniformes. Este movimiento es, en cierto sentido, el más "aleatorio" posible, y se le conoce como *distribución maxwelliana* —por el mismo James Clerk Maxwell que ya hemos encontrado antes—. Cuando el gas se halla en tal estado aleatorio, se dice que está en *equilibrio térmico.*

Existe un enorme volumen de puntos del espacio de fases que corresponden al equilibrio térmico.

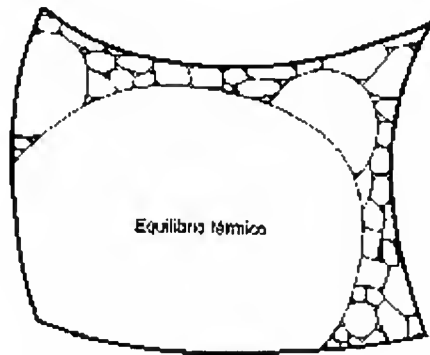


FIGURA VII.3. *Una división de grano-grueso del espacio de fases en regiones correspondientes a estados que son macroscópicamente indistinguibles uno de otro. La* entropía *es proporcional al logaritmo del volumen del espacio de fases.*

Los puntos de este volumen describen todos los detalles sobre las posiciones y las velocidades de las partículas individuales que son consistentes con el equilibrio térmico. Ese vasto volumen es uno de nuestros compartimentos del espacio de fases, el más grande y el que ocupa casi la totalidad del espacio de fases.

Consideremos otro posible estado del gas. Por ejemplo, aquel en el que todo el gas está concentrado en una esquina de la caja. De nuevo habrá muchos estados detallados individuales distintos, todos los cuales describen el gas concentrado de la misma forma. Son

macroscópicamente indistinguibles uno de otro, y los puntos del espacio de fases que los representan constituyen otro compartimento del espacio de fases. Sin embargo, el volumen de este compartimento resulta ser muchísimo más pequeño que el de los estados que representan el equilibrio térmico (en un factor de alrededor de $10^{10^{25}}$, si consideramos una caja de un metro cúbico que contiene aire a temperatura ambiente y presión atmosférica en equilibrio y si tomamos la región en la esquina como de un centímetro cúbico).

Para apreciar tales discrepancias entre volúmenes del espacio de fases, imaginemos una situación simplificada en la que cierto número de bolas van a distribuirse entre varias celdas. Supongamos que cada celda está vacía o contiene una sola bola. Las bolas representarán las moléculas del gas, y las celdas, las diferentes posiciones que podrían ocupar las moléculas en la caja. Seleccionemos un pequeño subconjunto de celdas como *especial;* éstas representarán las posiciones de las moléculas del gas que corresponden a la región en la esquina de la caja. Supongamos, para ser más precisos, que exactamente una décima parte de las celdas son especiales; digamos que hay *n* celdas especiales y *9n* no especiales (véase fig. VII.4).

Queremos distribuir *m* bolas entre las celdas de una manera aleatoria y encontrar la probabilidad de que todas ellas estén en celdas especiales. Si hay sólo una bola y diez celdas (de modo que tenemos una celda especial), esta probabilidad es evidentemente de un décimo. Lo mismo es válido si hay una bola y cualquier número 10*n* de celdas (con *n* celdas especiales). Así, para un "gas" con un solo átomo, el compartimento especial que corresponde al gas "concentrado en la esquina" tendrá un volumen de sólo un *décimo* del volumen total del "espacio de fases". Pero si aumentamos el número de bolas, la probabilidad de encontrarlas *todas* en celdas especiales disminuye. Para dos bolas, y digamos veinte celdas* (de las cuales dos son especiales) *(m*=2, *n*=2), la probabilidad es 1/190, o con cien celdas (con diez especiales) *(m*= 2, *n* =10) es de 1/110. Con un número muy grande de celdas, la probabilidad se hace 1/100.

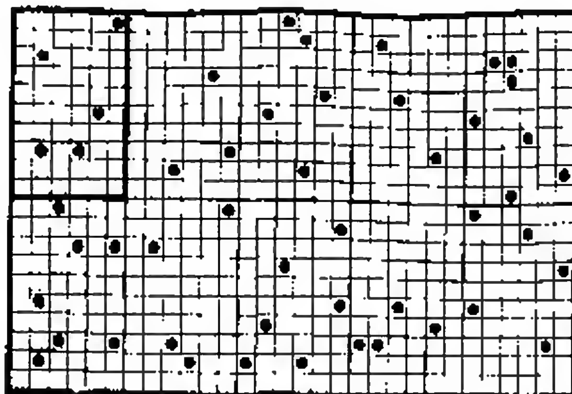


FIGURA VII.4. *Modelo para un gas en una caja: cierto número de minúsculas bolas se distribuye entre un número mucho mayor de celdas. Una décima parte de las celdas está etiquetada como* especiales. *Estas son las recuadradas en la esquina superior izquierda.*

---

* Para cualquier *m y n* la probabilidad es ${}^{10n}C_m \div {}^{10n}C_m = \dfrac{(10n)!(n-m)!}{n!(10n-m)!}$

Así, el volumen del compartimento especial para un "gas" de dos átomos es sólo una *centésima* parte del volumen total del "espacio de fases". Para *tres* bolas y treinta celdas (m = 3, n = 3), es 1/4060; y con un número muy grande de celdas se hace 1/1000, de modo que para un "gas" de tres átomos, el volumen del compartimento especial es de una *milésima* parte del volumen del "espacio de fases". Para cuatro bolas y un número muy grande de celdas, la probabilidad se hace 1/10000. Para cinco bolas y un número muy grande de celdas, la probabilidad se hace 1/100000, y así sucesivamente. Para m bolas y un número muy grande de celdas, la probabilidad se hace $1/10^m$, de modo que para un "gas" de m átomos, el volumen de la región especial es $1/10^m$ del volumen del "espacio de fases". (Esto sigue siendo cierto si se incluye el "momento".)

Podemos aplicar esto a un gas real en una caja, pero que, en lugar de ser sólo una décima parte del total, la región especial ocupa sólo una millonésima (esto es, 1/1000000) de este total (es decir, un centímetro cúbico en un metro cúbico). Esto significa que en lugar de ser la probabilidad $1/10^m$, ahora es $1/(1000000)^m$, esto es, $1/10^{6m}$. Para el aire ordinario habrá unas 10^{25} moléculas en total, así que tomamos $m = 10^{25}$. Por consiguiente, el compartimento especial del espacio de fases, que representa la situación en la que todo el gas está concentrado en la esquina, tiene un volumen de sólo

$$1/10^{60\,000\,000\,000\,000\,000\,000\,000\,000}.$$

de todo el espacio de fases.

La *entropía* de un estado es una medida del volumen V del compartimento que contiene los puntos del espacio de fases que representan a dicho estado. En vista de las enormes diferencias entre estos volúmenes como antes señalamos, es bueno que no tomemos la entropía como proporcional a dicho volumen sino al *logaritmo* del volumen:

$$\text{entropía} = k \log V.$$

El tomar un logaritmo ayuda a hacer estos números más razonables. El logaritmo* de 10000000, por ejemplo, es sólo alrededor de 16. La cantidad k es una constante, llamada *constante de Boltzmann.* Su valor es de alrededor de 10^{-23} julios por grado Kelvin. Aquí la razón esencial para tomar un logaritmo es hacer de la entropía una cantidad aditiva para sistemas independientes. Así, para dos sistemas físicos independientes, la entropía total de los dos sistemas combinados será la suma de las entropías de cada uno de ellos por separado. (Esto es una consecuencia de la propiedad algebraica elemental de la función logarítmica: $\log AB = \log A + \log B$. Si los dos sistemas pertenecen a compartimentos de volumen A y B, en sus respectivos espacios de fases, entonces el volumen del espacio de fases para los dos juntos será su producto AB, porque cada posibilidad para un sistema debe contarse una vez por cada posibilidad del otro. Por lo tanto, la entropía del sistema combinado es igual a la suma de las dos entropías individuales.)

Las enormes discrepancias entre los tamaños de los compartimentos del espacio de fases se ven más razonables en términos de entropía. La entropía de nuestra caja de gas de un metro cúbico de tamaño, como se describió antes, resultará ser del orden de unas 1400 JK^{-1} (= $14k \times 10^{25}$) veces mayor que la entropía del gas concentrado en la región especial de un centímetro cúbico de tamaño, puesto que $\log_e(10^{6x10^{25}})$ es alrededor de 14×10^{25}.

* El logaritmo utilizado aquí es un logaritmo *natural,* tomado con base e = 2.7182818285... en lugar de 10, pero esta diferencia apenas es importante. El logaritmo natural, $x = \log n$ de un número n es la potencia a la que debemos elevar e para obtener n, es decir, la solución de $e^x = n$.

Para dar los valores *reales* de la entropía para estos compartimentos tendríamos que ocuparnos un poco de la cuestión de las unidades a escoger (metros, julios, kilogramos, grados Kelvin, etc.). Eso estaría fuera de lugar aquí y, de hecho, para los enormes valores de la entropía que acabo de dar la elección particular de unidades no supone ninguna diferencia esencial. Sin embargo, a los expertos les advierto que tomaré unidades naturales, tal como las proporcionan las reglas de la mecánica cuántica, para las que la constante de Boltzmann resulta ser la *unidad:*

$$k=1.$$

LA SEGUNDA LEY EN ACCION

Supongamos, ahora, que partimos de un sistema en una situación muy especial, como sucedía cuando todo el gas estaba en un rincón de la caja. En los instantes siguientes el gas se expandirá y ocupará rápidamente volúmenes cada vez mayores. Tras cierto tiempo, se establecerá el equilibrio térmico. ¿Cuál es nuestra imagen de este proceso en términos del espacio de fases?

En cada paso, el estado detallado de posiciones y movimientos de las partículas vendrá descrito por un simple punto en el espacio de fases. A medida que el gas evoluciona, este punto se mueve por el espacio de fases y sus recorridos precisan toda la historia de las partículas del gas. El punto parte de una región pequeñísima, la región que representa la colección de posibles estados iniciales, para los cuales todo el gas está en una esquina particular de la caja.

A medida que el gas comienza a expandirse, nuestro punto en movimiento entrará en un volumen mayor del espacio de fases, que corresponde a los estados en los que el gas se ha expandido por la caja. El punto en el espacio de fases sigue invadiendo volúmenes mayores a medida que el gas se extiende, de manera que cada nuevo volumen empequeñece totalmente a aquellos en los que el punto había estado antes ¡por factores absolutamente extraordinarios! (véase fig. VII.5). En cada caso, una vez que el punto haya entrado en el volumen mayor, no existirá posibilidad de que halle los volúmenes anteriores más pequeños. Finalmente, se perderá en el volumen más grande de todos: el correspondiente al equilibrio térmico. Este volumen ocupa prácticamente la totalidad del espacio de fases.

Podemos asegurar virtualmente que, en su deambular aleatorio, nuestro punto en el espacio de fases no encontrará pronto ninguno de los volúmenes pequeños. Una vez que se haya alcanzado el estado de equilibrio térmico, el estado permanecerá allí, efectivamente, para siempre. Vemos así que la entropía del sistema, al proporcionar una medida logarítmica del volumen del compartimento apropiado del espacio de fases, tenderá inexorablemente a aumentar.*

¿Estamos ante una *explicación* de la segunda ley? En efecto, podemos suponer que nuestro punto en el espacio de fases no se mueve de ninguna

* Por supuesto no es cierto que nuestro punto en el espacio de fases no encontrará *nunca* de nuevo uno de los compartimentos más pequeños. Si esperamos el tiempo suficiente llegará a reentrar en estos volúmenes relativamente minúsculos. (Esto se conoce como *recurrencia de Poincaré.)* Sin embargo, las escalas de tiempo serían ridículamente largas en la mayoría de las circunstancias, *v.g.* alrededor de $10^{10^{26}}$ años para el caso del gas que se concentra en un centímetro cúbico en una esquina de la caja. Esto es muchísimo mayor que el tiempo de existencia del Universo. Descartaré esta posibilidad en la discusión que sigue, ya que no es realmente importante para el problema que consideramos

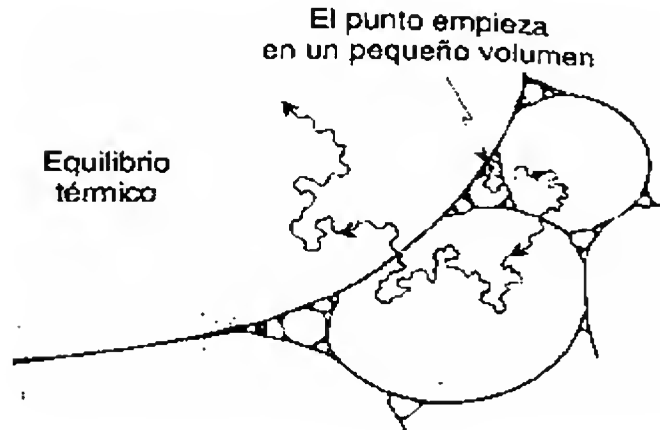


FIGURA VII.5. *La segunda ley de la termodinámica en acción: conforme avanza el tiempo, el punto en el espacio de fases invade gradualmente compartimientos de cada vez mayor volumen. Como resultado la entropía se incrementa continuamente.*

manera especialmente concebida, y si parte de un volumen pequeñísimo en el espacio de fases, correspondiente a una entropía *pequeña,* entonces a medida que el tiempo avanza será extraordinariamente probable que entre en volúmenes del espacio de fases sucesivamente mayores (los cuales corresponderán a valores gradualmente crecientes de la entropía).

Pero hay algo extraño en lo que se sigue a partir de este argumento: parece que hemos deducido una conclusión con *asimetría temporal.* La entropía *crece* en la dirección positiva del tiempo, y por ello debería *disminuir* en la dirección inversa. ¿De dónde procede tal asimetría temporal? No hemos introducido ninguna ley física con asimetría temporal. La asimetría temporal procede simplemente del hecho de que el sistema ha partido de un estado muy especial (esto es, de baja entropía), y al haber empezado así lo hemos visto evolucionar en la dirección del *futuro* y hemos encontrado que la entropía crece. Este incremento de la entropía va de acuerdo con el comportamiento de los sistemas en nuestro universo, pero podríamos perfectamente haber aplicado este mismo argumento en la dirección inversa del tiempo. Especificaremos de nuevo que en un instante dado el sistema se halla en un estado de baja entropía. ¿Cuál es la secuencia más probable de estados que lo preceden?

Ensayemos el argumento en forma inversa. Consideremos, como antes, que el estado de baja entropía es el de todo el gas en una esquina de la caja. Nuestro punto en el espacio de fases está ahora en la misma región minúscula de la que partimos antes, pero sigamos su historia *hacia atrás.* Si imaginamos que el punto en el espacio de fases deambula de una forma aleatoria, como antes, entonces esperamos que, a medida que seguimos este movimiento hacia atrás en el tiempo, alcanzará pronto el volumen considerablemente mayor del espacio de fases que alcanzó antes, correspondiente al gas extendido un poco por la caja, pero no en equilibrio térmico. Y seguirá alcanzando volúmenes cada vez mayores, a tal grado que cada nuevo volumen empequeñecerá a los previos. Y así, si retrocedemos todavía más en el tiempo, lo encontraremos en el volumen más grande de todos: el equilibrio térmico.

Ahora hemos deducido que en cierto instante todo el gas estaba concentrado en la esquina de la caja. Partió del equilibrio térmico, luego comenzó a concentrarse en un extremo de la caja, más y más, hasta que se agrupó en el pequeño volumen especificado de la esquina. La entropía tendría que haber estado *decreciendo* todo el tiempo: partió del valor alto de equilibrio y luego decreció

gradualmente hasta que alcanzó el valor muy bajo correspondiente al gas concentrado en la pequeña esquina de la caja.

Por supuesto, esto no se parece en nada a lo que realmente sucede en nuestro universo. La entropía no sólo no disminuye, sino que *aumenta.* Si supiéramos que todo el gas estaba concentrado en un rincón de la caja en un instante particular, entonces una situación mucho más probable que hubiera precedido a aquélla habría sido la de que el gas se hubiera mantenido en el rincón mediante una partición, que hubiese sido rápidamente eliminada. O quizá el gas estuvo en un estado congelado o líquido y fue calentado rápidamente para volverse gaseoso. Para cualquiera de estas posibles opciones, la entropía era aún *menor* en los estados previos. La segunda ley realmente se satisfacía y la entropía estaba creciendo, dicho de otra forma, en la dirección *inversa* del tiempo, realmente *disminuía.*

Ahora tenemos que nuestro argumento nos ha dado la respuesta equivocada. Nos ha dicho que la forma más probable de tener el gas en la esquina de la caja sería partir del equilibrio térmico y que, con la entropía reduciéndose continuamente, el gas se agruparía en la esquina. Mientras que, de hecho, es extraordinariamente improbable que suceda así en nuestro mundo real. En éste, el gas partiría de un estado *aún* menos probable (esto es, de menor entropía), y la entropía aumentaría continuamente hasta alcanzar el valor que tendrá posteriormente cuando el gas esté concentrado en la esquina.

Nuestro argumento parece ser bueno cuando se aplica en la dirección del futuro, pero no en la dirección del pasado. Para el futuro prevemos correctamente que, de donde quiera que parta el gas en la esquina, lo más probable es que llegue a alcanzar el equilibrio térmico, y *no* que aparezca una partición o que el gas se congele o se haga líquido. Semejantes posibilidades representarían el tipo de comportamiento con disminución de entropía en la dirección del futuro que ya hemos descartado. En la dirección del pasado, en cambio, semejantes opciones "anómalas" son

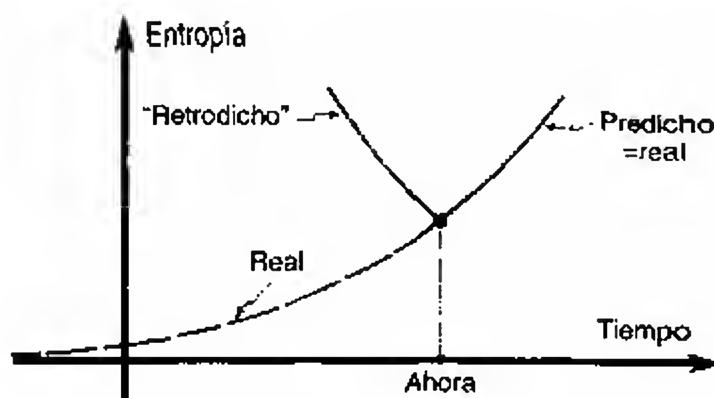


FIGURA VII.6. *Si aplicamos el argumento representado en la fig. VIL 5 pero con una dirección inversa del tiempo, "retrodecimos"* que la entropía debe también incrementarse en el* pasado, *a partir de su valor presente. Esto está en abierta contradicción con la observación.*

* Así como "predicción" implica referencia al futuro, el autor acuña el término "retrodicción" para indicar un "supuesto con relación al pasado". [N. del E.]

las que sucederían —y no serían en absoluto anómalas. Nuestro argumento del espacio de fases nos dio una respuesta totalmente errónea cuando tratamos de aplicarla en la duración inversa del tiempo.

Evidentemente, todo lo anterior arroja algunas dudas sobre el argumento original. *No* hemos deducido la segunda ley. Lo que realmente mostraba el argumento era que para un estado de baja entropía dado (digamos para un gas concentrado en una esquina de una caja), entonces, *en ausencia de otros factores que impongan restricciones al sistema,* se espera que la entropía aumente en *ambas* direcciones en el tiempo a partir del estado dado (véase fig. VII.6). Y si tal argumento no ha funcionado en la dirección del pasado, es precisamente porque *existían* tales factores: había algo en el pasado que limitaba al sistema, algo que *obligaba* a la entropía a ser baja. La tendencia hacia una entropía alta en el futuro no es sorpresa. Los estados de alta entropía son, en cierto sentido, los estados "naturales" que no necesitan más explicación, pero los estados de baja entropía en el pasado son un enigma. ¿Qué obligaba a la entropía de nuestro mundo a ser tan baja en el pasado? La presencia común de estados para los que la entropía es absurdamente baja, es un hecho sorprendente del universo real en el que habitamos —aunque tales estados son tan comunes y familiares para nosotros que no tendemos a sorprendernos con ellos—. Nosotros mismos somos configuraciones de entropía ínfima. No deberíamos entonces alarmarnos si, *dado* un estado de baja entropía, la entropía después resulta ser mayor. Lo que *debería* sorprendernos es que la entropía se haga más pequeña a medida que nos adentramos en el pasado.

EL ORIGEN DE LA BAJA ENTROPÍA EN EL UNIVERSO

Trataremos de entender de dónde procede esta "sorprendentemente" baja entropía en que *se* halla el mundo real que habitamos. Empecemos por nosotros mismos. Si podemos entender de dónde procede nuestra propia baja entropía, entonces seríamos capaces de ver *de* dónde procede la baja entropía del gas mantenido por la partición o del vaso de agua en la mesa, o del huevo mantenido sobre la sartén, o del terrón de *azúcar* mantenido sobre la taza de café. En cada caso, una persona o un grupo de personas (o quizá una gallina) era responsable directa o indirectamente. Era, en gran medida, una pequeña parte de nuestra propia baja entropía la que se utilizaba para establecer estos otros estados de baja entropía, pero podrían haber estado involucrados otros factores: quizá se utilizó una bomba de vacío para aspirar el gas hasta la esquina de la caja tras la partición o, si la bomba no se operó manualmente, pudo haberse quemado algún combustible fósil *(v. gr.* aceite) para obtener la energía de baja entropía indispensable para su operación. Tal vez la bomba fue operada eléctricamente y dependió, entonces de la energía de baja entropía almacenada en combustible de uranio de una estación nuclear. Más adelante regresaré a estas otras fuentes de baja entropía, pero antes consideremos simplemente la baja entropía en nosotros mismos.

¿De dónde procede nuestra propia baja entropía? La organización de nuestros cuerpos es tal debido al alimento que comemos y al oxígeno que respiramos. Con frecuencia se oye decir que obtenemos *energía* de nuestra ingestión de alimentos y oxígeno, pero hay un sentido evidente en el que esto no es correcto. Es cierto que el alimento que consumimos se combina con el oxígeno que introducimos en nuestros cuerpos, y que esto nos proporciona energía, pero esa energía, en su mayor parte, escapa de nuevo de nuestros cuerpos, principalmente en forma de calor.

Puesto que la energía se conserva, y puesto que el contenido real de energía de nuestros cuerpos permanece más o menos constante a lo largo de nuestra vida, no hay necesidad de *añadir* nada al contenido de energía de nuestros cuerpos: no *necesitamos* más energía de la que ya tenemos. Añadimos algo a nuestro contenido energético cuando aumentamos de peso, pero normalmente esto no se considera deseable. Cuando somos niños, crecemos y también así aumentamos nuestro contenido energético, pero ahora no es esto en lo que estoy interesado. La cuestión es cómo nos mantenemos *vivos* a lo largo de nuestra vida normal (principalmente adulta). Para esto *no* necesitamos añadir nada a nuestro contenido energético.

Lo que sí necesitamos es reemplazar la energía que perdemos continuamente en forma de calor. En realidad, cuanto más "energéticos" somos, más energía perdemos, y toda esa energía debe ser reemplazada.

El calor es la forma más *desordenada* de energía que existe, la forma con mayor entropía. Tomamos energía en forma de *baja* entropía *(alimento* y *oxígeno*) y la desechamos en una forma de *alta* entropía (calor dióxido de carbono, excrementos). No necesitamos obtener energía dé nuestro medio ambiente porque la energía se *conserva,* pero estamos luchando continuamente contra la segunda ley de la termodinámica. La entropía *no* se conserva, *aumenta* todo el tiempo. Para mantenernos vivos necesitamos reducir la entropía en nosotros. Lo hacemos alimentándonos con una combinación (de baja entropía) de comida y oxígeno atmosférico, combinándolos dentro de nuestro cuerpo, y desechando la energía, que de no hacerlo hubiéramos ganado, en una forma de alta entropía. De este modo evitamos que aumente la entropía en nuestros cuerpos, y podemos mantener e incluso incrementar nuestra organización interna. (Véase Schrödinger, 1967.)

¿De dónde procede el suministro de baja entropía? Si el alimento que estamos comiendo es carne (o setas), entonces *este alimento,* al igual que nosotros, tendría que depender de una fuente externa adicional de baja entropía que proporcione y mantenga su estructura de baja entropía. Esto simplemente desplaza a otra parte el problema del origen de la baja entropía externa. Supongamos que nosotros (o el animal, o la seta) estamos consumiendo una *planta.* Debemos estar todos profundamente agradecidos (directa o indirectamente) a las plantas verdes por su "inteligencia": tomar dióxido de carbono atmosférico, separar el oxígeno del carbono y utilizar el carbono para formar su propia sustancia. Este procedimiento, la *fotosíntesis,* reduce considerablemente la entropía. Nosotros mismos hacemos uso efectivo de esta separación de baja entropía mediante la recombinación simple del oxígeno y el carbono dentro de nuestros cuerpos. ¿Cómo es que las plantas verdes son capaces de conseguir esta mágica reducción de entropía? Lo hacen utilizando la *luz del Sol.*

La luz que procede del Sol trae energía a la Tierra, en una forma de comparativamente *baja* entropía: en los fotones de la luz visible. La Tierra, incluso sus habitantes, no *retiene* esta energía sino que (después de un rato) la reirradia toda hacia el espacio. Sin embargo, la energía reirradiada está en una forma de alta entropía, que se llama "calor radiante" (lo que significa fotones infrarrojos). Contrariamente a la impresión común, la Tierra (junto con sus habitantes) *no gana* energía del Sol. Lo que hace la Tierra es tomar energía en una forma de baja entropía y luego arrojarla *toda* al espacio, pero en una forma de alta entropía (fig. VII.7). Lo que el Sol ha hecho por nosotros es suministrarnos una fuente enorme de baja entropía y nosotros (por vía de la "inteligencia" de las

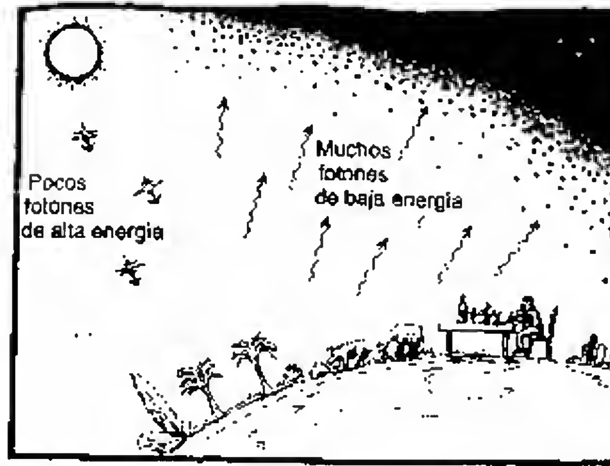


FIGURA VII.7. *Cómo hacemos uso del hecho de que el Sol sea un punto caliente en la oscuridad del espacio.*

plantas), hacemos uso de ésta, extrayendo finalmente una mínima parte de su baja entropía empleándola para hacer posibles estas intrincadas estructuras organizadas que somos nosotros mismos.

Veamos, desde el punto de vista global con relación al Sol y la Tierra, qué ha sucedido con la energía y la entropía. El Sol emite energía en forma de fotones de luz visible. Algunos de estos fotones son absorbidos por la Tierra y su energía es reirradiada en forma de fotones infrarrojos. Ahora bien, la diferencia crucial entre los fotones de luz visible y los infrarrojos es que los primeros tienen una frecuencia mayor y, por lo tanto, individualmente tienen más energía que los últimos. (Recordemos la fórmula de Planck $E = h\nu$, dada: cuanto mayor sea la frecuencia de un fotón, mayor será su energía.) Puesto que cada uno de los fotones de luz visible tiene mayor energía que cada uno de los infrarrojos, los fotones de luz visible que llegan a la Tierra deben hacerlo en número *menor* al de los fotones infrarrojos que la dejan, de modo que la *energía* que entra en la Tierra compensa la que la abandona. La energía que la Tierra arroja al espacio se distribuye sobre muchos más grados de libertad que la energía que se recibe del Sol. Puesto que hay muchos más grados de libertad involucrados cuando la energía es devuelta, el volumen del espacio de fases es mayor y la *entropía* aumenta enormemente. Las plantas verdes, al tomar energía bajo la forma de baja entropía (comparativamente *pocos* fotones de luz visible) y reradiarla en la forma de alta entropía (comparativamente muchos fotones infrarrojos), han sido capaces de alimentarse de ella y proporcionarnos la separación oxígeno-carbono que requerimos.

Todo esto es posible por el hecho de que el Sol es un *punto caliente* en el cielo. El cielo tiene una temperatura desigual: una pequeña región, la ocupada por el Sol, está a una temperatura mucho más alta que el resto tal hecho nos proporciona entonces baja entropía. La Tierra obtiene energía de este punto caliente en forma de baja entropía (pocos fotones), y la reirradia a las regiones frías en forma de alta entropía (muchos fotones).

¿Por qué el Sol es ese punto caliente? ¿Cómo ha podido conseguir esa desigualdad de temperatura y proporcionarnos así un estado de baja entropía? La respuesta es que se ha formado por contracción gravitatoria a partir de una previa distribución uniforme de gas (principalmente hidrógeno). A medida que se contrajo, en los primeros pasos de su formación, el Sol se calentó. Seguiría contrayéndose y calentándose aún más si no fuera porque, cuando su temperatura y su

presión alcanzan un cierto punto, encuentra otra fuente de energía —además de la contracción gravitatoria— en las *reacciones termonucleares:* la fusión de núcleos de hidrógeno en núcleos de helio para dar energía. Sin la reacciones termonucleares, el Sol se hubiera hecho mucho *más caliente* y más pequeño de lo que es ahora, hasta que finalmente hubiera muerto. Las reacciones nucleares han impedido que el Sol se vuelva *demasiado* caliente, deteniendo su contracción, y han estabilizado al Sol a una temperatura que es apropiada para nosotros, haciendo posible que siga brillando durante un tiempo mucho mayor de lo que lo hubiera hecho en otro caso.

Es necesario hacer notar que, aunque las reacciones termonucleares son importantes para determinar la naturaleza y la cantidad de la energía radiada por el Sol, lo crucial es la *gravitación.* (La potencialidad para las reacciones termonucleares *da* una contribución altamente significativa al bajo valor de la entropía solar, pero las cuestiones que plantea la entropía de la fusión son delicadas y una completa discusión de ellas sólo serviría para complicar el argumento sin afectar la conclusión final.)[3] Sin gravedad ni siquiera existiría el Sol. El Sol brillaría aún sin reacciones termonucleares —aunque no de una forma apropiada para nosotros__, pero no estaría brillando en absoluto sin la gravedad que se necesita para mantenerlo unido y proporcionar las temperaturas y presiones idóneas. Sin gravedad, todo lo que tendríamos sería un gas frío y difuso en el lugar del Sol y *no* habría punto caliente en el cielo.

No he discutido todavía la fuente de la baja entropía en los "combustibles fósiles" de la Tierra, pero las consideraciones son básicamente las mismas. Según la teoría convencional, todo el petróleo (y gas natural) procede de la vida vegetal prehistórica. De nuevo son las plantas las que resultan haber sido responsables de esta fuente de baja entropía. Las plantas prehistóricas obtuvieron su energía del Sol, de modo que debemos volver otra vez a la acción gravitatoria que formó al Sol a partir de un gas difuso.

Hay una interesante teoría "inconformista" alternativa acerca del origen del petróleo en la Tierra, debida a Thomas Gold, que cuestiona esta opinión convencional sugiriendo que hay mucho más petróleo en la Tierra del que podría haber surgido a partir de plantas prehistóricas. Gold cree que el petróleo estaba atrapado en el interior de la Tierra cuando ésta se formó, y que desde entonces ha estado rezumando continuamente, acumulándose en bolsas subterráneas.[4] No obstante, según la teoría de Gold el petróleo habría sido en cualquier caso sintetizado por la luz del Sol, aunque esta vez en el espacio, antes de que la Tierra se formase. De nuevo sería el Sol —formado gravitatoriamente— el responsable.

¿Qué hay de la energía nuclear de baja entropía en el isótopo de uranio-235 que se utiliza en las centrales nucleares? Esta no procedía originalmente del Sol (aunque podría haber pasado

---

[3] Se gana entropía al combinar núcleos ligeros *(v.g.* de hidrógeno) en las estrellas para dar núcleos más pesados (de helio o, en última instancia, de hierro). Análogamente, existe baja entropía en el hidrógeno presente en la Tierra, parte de la cual podemos utilizar convirtiendo el hidrógeno en helio en las centrales nucleares de "fusión". La posibilidad de ganar entropía por estos medios aparece únicamente debido a que la gravitación ha hecho posible que los núcleos se concentren, al margen de los fotones mucho más numerosos que han escapado a la inmensidad del espacio y constituyen la radiación de fondo de cuerpo negro de 2.7 K . Tal radiación contiene una entropía mucho mayor que la de la materia de las estrellas ordinarias, y si se concentrara de nuevo en el material de las estrellas servirla para desintegrar otra vez la mayor parte de los pesados núcleos en sus partículas constituyentes. Por ello, la ganancia de entropía en la fusión es temporal, y se hace posible sólo mediante los efectos concentrantes de la gravedad. Veremos más adelante que incluso aunque la entropía disponible por la vía de la fusión de los núcleos sea muy grande en relación con la obtenida a través de la gravedad —y la entropía en la radiación de fondo es enormemente mayor—, éste será un estado de cosas únicamente local y temporal. Las reservas de entropía de la gravitación son *enormes* en comparación con las de la fusión o las de la radiación de 2.7 K .

[4] Alguna evidencia reciente a partir de los pozos ultraprofundos perforados en Suecia puede interpretarse como un apoyo a la teoría de Gold, pero el tema es controvertido y existen otras explicaciones convencionales.

perfectamente por el Sol en alguna etapa) sino de alguna otra estrella que brotó hace muchos miles de millones de años a partir de una explosión de supernova. En realidad, el material fue recogido de *muchas* de estas estrellas explosivas y parte de él se juntó finalmente (mediante la actuación del Sol) para proporcionar los elementos pesados en la Tierra, incluso el uranio-235: cada núcleo atómico, con su reserva de energía de baja entropía, procede de los violentos procesos nucleares que tuvieron lugar durante la explosión de la supernova.

La explosión ocurrió a consecuencia del colapso gravitatorio[5] de una estrella cuya masa era demasiado grande para poder ser mantenida por las fuerzas debidas a la presión térmica. Como resultado del colapso, y la subsiguiente explosión, quedó un pequeño núcleo, probablemente en la forma de lo que se conoce como una *estrella de neutrones* (volveré a esto más adelante.) La estrella se habría contraído gravitatoriamente en su origen a partir de una difusa nube de gas y mucho de ese material original, incluido el uranio-235, habría sido arrojado al espacio. No obstante, hubo una enorme ganancia en entropía debida a la contracción gravitatoria, a causa del núcleo de la estrella de neutrones que quedó. De nuevo fue la *gravedad* la responsable última; esta vez al causar la condensación (finalmente violenta) de gas difuso en una estrella de neutrones.

Parece que hemos llegado a la conclusión de que toda esta notable pequeñez de la entropía que encontramos —y que proporciona el aspecto más enigmático de la segunda ley— debe atribuirse al hecho de que se pueden ganar grandes cantidades de entropía mediante la contracción gravitatoria de gas difuso en estrellas. ¿De dónde procede todo ese gas? Es precisamente el hecho de que este gas que empezó siendo *difuso* el que nos proporciona una enorme reserva de baja entropía. Aún estamos viviendo de esa baja entropía, y continuaremos haciéndolo durante mucho tiempo. Es la potencialidad de agrupamiento gravitatorio de este gas la que nos ha dado la segunda ley; además, no es solamente la segunda ley lo que este agrupamiento gravitatorio ha producido, sino algo mucho más preciso y detallado que el simple enunciado: "La entropía del mundo empezó siendo muy baja". La entropía podía habernos sido dada "baja" de muchas *otras* formas diferentes, es decir, podría haber habido mucho "orden manifiesto" en el Universo primitivo, pero muy diferente del que se nos presenta en el mundo real. (Imaginemos que el Universo hubiera sido un dodecaedro regular —como le hubiera gustado a Platón— o alguna otra forma geométrica improbable. Esto sería realmente "orden manifiesto", pero no del tipo que esperamos encontrar en el Universo primitivo *real)* Debemos comprender de dónde procede todo este gas difuso, y para ello tendremos que volver a nuestras teorías cosmológicas.

LA COSMOLOGIA Y EL BIG BANG O GRAN EXPLOSIÓN

Hasta donde nos permiten asegurarlo nuestros más potentes telescopios (tanto los ópticos como los de radio), el Universo es uniforme a gran escala, y se encuentra en *expansión.* Cuanto más lejos miramos, más rápido se alejan de nosotros los cuásares y las galaxias. Es como si el Universo mismo se hubiera creado a partir de una gigantesca explosión, un suceso conocido como el *big bang,* que ocurrió hace unos diez mil millones de años.*

---

[5] Estoy suponiendo que se trata de lo que se conoce como una supernova de "tipo II". Si hubiera sido una supernova de "tipo I" estaríamos pensando de nuevo en términos de la ganancia "temporal" de entropía proporcionada por la "fusión" *(cfr.* nota 3). Es poco probable que las supernovas tipo I produzcan mucho uranio.

* Hay una enconada discusión a propósito del valor de esta cifra, que está entre 6×10^9 y 1.5×10^{10} años. Se trata de valores

Un impresionante apoyo adicional para esta uniformidad, y para la existencia real del *big bang,* viene de lo que se conoce como *radiación de fondo de cuerpo negro.* Esta es radiación térmica —fotones que se mueven aleatoriamente, sin ninguna fuente discernible— correspondiente a una temperatura de aproximadamente 2.7° absolutos (2.7 K), esto es -270.3° Celsius, o 454.5° Fahrenheit bajo cero. Puede parecer una temperatura *muy* fría —como de hecho lo es—, pero es el residuo del "flash" del propio *big bang.*

Puesto que el Universo se ha expandido en un factor tan enorme desde ese momento, esta bola de fuego inicial debe haberse extendido multiplicándose por un factor enorme. Las temperaturas durante la gran explosión excedían cualquiera que pudiera darse en el tiempo presente, pero —debido a la expansión— esta temperatura se ha enfriado hasta el pequeñísimo valor que tiene ahora la radiación negra de fondo. La presencia de esta radiación de fondo fue *predicha* por el físico ruso-estadounidense George Gamow en 1948, sobre la base de la imagen del *big bang* ahora estándar, y fue observada accidentalmente por Penzias y Wilson en 1965.

Abordaré ahora una cuestión desconcertante: si las galaxias distantes están alejándose de nosotros, ¿no significa eso que nosotros mismos estamos ocupando algún lugar central muy especial? ¡Nada de eso! La misma recesión de galaxias distantes se vería *donde quiera* que pudiéramos estar situados en el Universo. La expansión es uniforme a gran escala, y ninguna localización particular es preferida a otra. Esto se representa a menudo en términos de un globo que se infla (fig. VII.8). Supongamos que hay manchas en el globo que representan las diferentes galaxias, y tomemos la propia superficie tridimensional del globo para representar el universo tridimensional completo. Es evidente que con referencia a *cada* punto del globo, todos los demás puntos se alejan de él. No hay ningún punto del globo preferido. E igual, como se ve a partir desde la posición de cada galaxia en el Universo, todas las demás galaxias parecen alejarse de ella en todas direcciones. Este globo en expansión proporciona una imagen bastante buena de

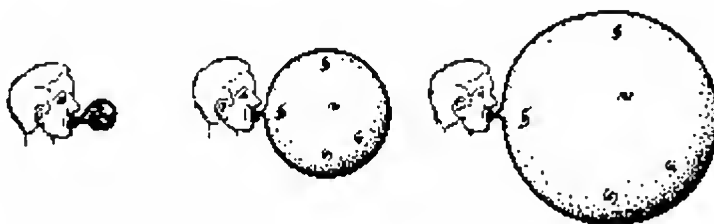


FIGURA VII.8. *Se puede hacer una analogía entre la expansión del Universo y la superficie de un globo que se hincha. Todas las galaxias se alejan una de otra.*

uno de los tres modelos de universo estándar del tipo llamado de *Fríedman-Robertson-Walker* (FRW), a saber: el modelo FRW con curvatura positiva espacialmente cerrado. En los otros dos modelos-FRW (con curvatura nula o negativa), el Universo se expande de un modo semejante, pero en lugar de tener un universo finito, como indica la superficie del globo, tenemos un universo infinito con un número infinito de galaxias.

En el más fácil de comprender de estos dos modelos infinitos, la geometría es *euclidiana,* esto es, tiene curvatura *nula.* Pensemos en un plano ordinario, que representa el universo espacial

considerablemente mayores que los 10^9 años que parecieron adecuados después de que Edwin Hubble, alrededor de 1930, demostrara que el Universo se expande.

completo en donde marcamos puntos que representan a las galaxias. A medida que el Universo evoluciona, esas galaxias se alejan unas de otras de manera uniforme. Pensemos en esto en términos de *espacio-tiempo:* tenemos un plano euclidiano distinto para cada "instante de tiempo", y todos esos planos se conciben como apilados uno sobre otro, de modo que tenemos una imagen de todo el espacio-tiempo a la vez (fig. VII.9). Las galaxias se representan ahora como curvas —las líneas de universo de las historias de las galaxias— y estas curvas divergen en la dirección del futuro. De nuevo, no existe ninguna línea de universo de una galaxia preferida.

Para el modelo FRW restante, el modelo con curvatura *negativa,* la geometría espacial es la geometría de Lobachevsky *no* euclidiana descrita en el capítulo V e ilustrada con el grabado de Escher en la fig. V.2. Para la descripción del espacio-tiempo necesitamos uno de estos espacios de Lobachevsky por cada "instante de tiempo", y apilamos todos ellos uno encima de otro para dar una imagen del espacio-tiempo completo (fig. VII.10).[6] Una vez más, las líneas-de-universo de las galaxias son curvas que se separan unas de otras en dirección al futuro, y no hay ninguna galaxia preferida.

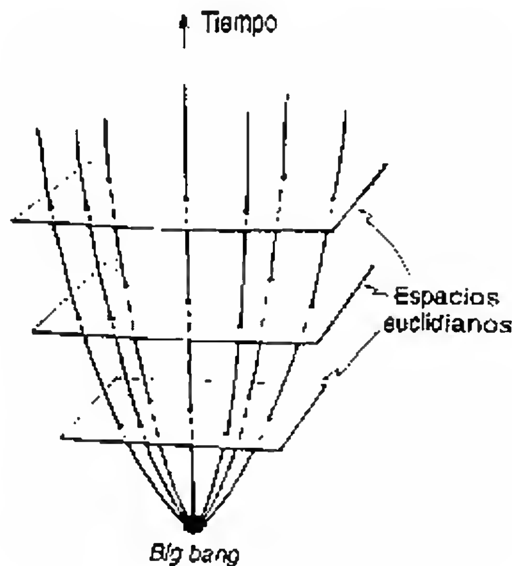


FIGURA VII.9. *Representación espacio-temporal de un universo en expansión con secciones espaciales euclidianas (se muestran dos dimensiones espaciales).*

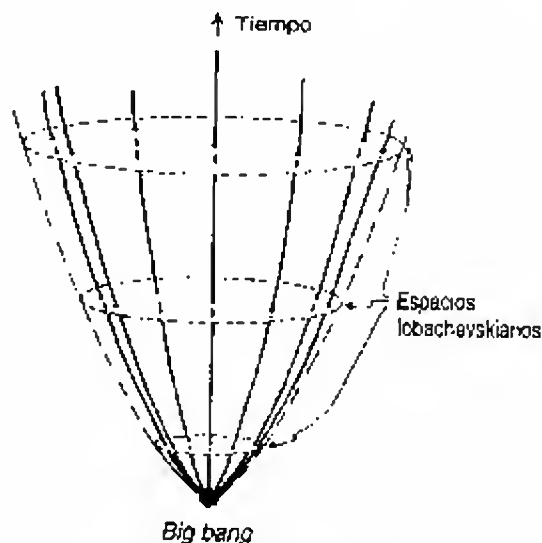
Por supuesto, en mis descripciones he suprimido una de las tres dimensiones espaciales (como hicimos en el capítulo V), a fin de dar un espacio-tiempo tridimensional más cercano a lo que requeriría la imagen completa del espacio-tiempo tetradimensional.

Aún así, es difícil visualizar el espacio-tiempo con curvatura positiva sin descartar una dimensión espacial más. Hagámoslo y representemos el Universo espacialmente cerrado con curvatura positiva mediante un *círculo* (unidimensional), en lugar de la esfera (bidimensional) que había sido la superficie del globo. A medida que el universo se expande, ese *círculo* aumenta

[6] He llamado a los modelos con curvatura nula o negativa modelos *infinitos.* Existen no obstante, maneras de "plegar" estos modelos para hacerlos finitos. Tal consideración —de no poco probable importancia para el universo real— no afectaría en gran medida a la discusión.

de tamaño y podemos representar el espacio-tiempo apilando estos círculos (un círculo para cada "instante de tiempo") uno encima de otro para obtener una especie de cono curvo (fig. VII.11a).

FIGURA VII 10. *Representación espacio-temporal de un universo en expansión con secciones espaciales lobachevskianas (se muestran dos dimensiones espaciales).*



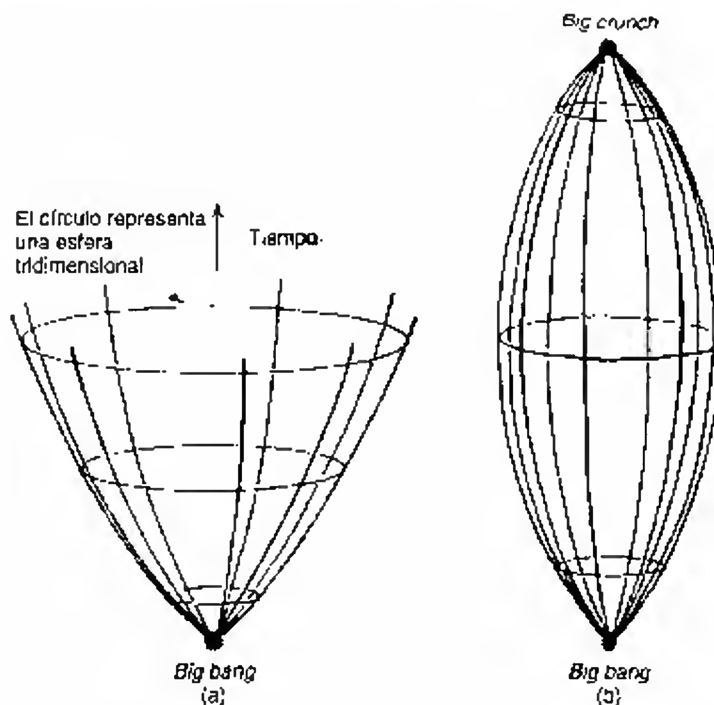


FIGURA VII. 11. *(a) Representación espacio-temporal de un universo en expansión con secciones espaciales esféricas; sólo se muestra una dimensión, (b) Este universo puede llegar a recolapsar en un* big crunch *final.*

Ahora bien, de las ecuaciones de la relatividad general de Einstein se deduce que este universo cerrado no puede seguir expandiéndose siempre. Después de alcanzar una etapa de máxima expansión, colapsa sobre sí mismo para llegar finalmente otra vez al tamaño nulo en una especie de *big bang* al revés (fig. VII.11b). Este big bang invertido en el tiempo se denomina a veces el

big crunch o gran colapso. Los modelos FRW de universo (infinito), con curvatura negativa y nula, no colapsan de esta forma. En lugar de llegar a un gran colapso, continúan expandiéndose siempre.

Esto es cierto al menos para la relatividad general *estándar* en la que la llamada *constante cosmológica* se considera nula. Con valores no nulos apropiados para la constante cosmológica, es posible tener modelos de universo espacialmente infinitos que colapsen a un *big crunch,* o modelos finitos con curvatura positiva que se expandan indefinidamente. La presencia de una constante cosmológica no nula complicaría ligeramente esta idea, aunque no de manera significativa para nuestros propósitos. Por simplicidad, consideraré nula la constante cosmológica.* (En el momento de escribir este libro se sabe, a partir de las observaciones, que la constante cosmológica debe ser muy pequeña, y los datos son consistentes con que sea nula. Para más información sobre los modelos cosmológicos, véase Rindler, 1977.)

Desafortunadamente los datos no son todavía suficientemente buenos como para apuntar claramente a uno u otro de los modelos cosmológicos propuestos (ni para determinar si la presencia o no de una pequeñísima constante cosmológica pudiera tener algún efecto global significativo). Por otra parte, los datos indican que el universo tiene curvatura espacial negativa (con geometría de Lobachevsky a gran escala) y que continuará expandiéndose indefinidamente. Esto se basa principalmente en observaciones sobre la cantidad de materia presente en forma visible. Sin embargo, puede haber enormes cantidades de materia invisible distribuida por todo el espacio, en cuyo caso el Universo podría estar curvado positivamente y podría colapsar finalmente en un *big crunch,* aunque en una escala de tiempo mucho mayor que la de los 10^{10} años, aproximadamente, que lleva existiendo el Universo. Para que tal colapso fuera posible, tendría que haber esparcida por el espacio, en esta forma invisible (la postulada "materia oscura"), una cantidad de materia unas treinta veces mayor que la que puede advertirse a través de los telescopios. Existe evidencia indirecta de que realmente hay presente una cantidad sustancial de materia oscura, pero es todavía una cuestión abierta el que sea o no suficiente para "cerrar el universo" (o hacerlo espacialmente plano) y colapsarlo.

LA BOLA DE FUEGO PRIMORDIAL

Volvamos a nuestra búsqueda del origen de la segunda ley de la termodinámica, el cual hemos rastreado hasta la presencia del gas difuso a partir del cual se condensaron las estrellas. ¿Qué es este gas? ¿De dónde procede? En su mayor parte es hidrógeno, pero también hay alrededor de 23% (en masa) de helio y pequeñas cantidades de otros materiales. Según la teoría estándar, este gas fue "escupido" como consecuencia de la explosión que creó el Universo: el *big bang.* Sin embargo, no fue esa una explosión ordinaria en la que el material se expele a partir de un punto central hacia el espacio preexistente: aquí, el propio espacio es *creado* en la explosión y no existe, ni existió, punto central.

Quizá sea más fácil visualizar la situación en el caso de curvatura positiva. Consideremos de nuevo la fig. VII. 11, o el globo inflado de la fig. VII.8. No hay "espacio vacío preexistente" en el cual se vacíe la materia producida por la explosión; el mismo espacio, es decir, la "superficie del globo", nace de la explosión. Debemos darnos cuenta de que, si bien en nuestras

* Einstein introdujo la constante cosmológica en 1917, pero se retractó hacia 1931, al considerar su introducción anterior como ¡su "mayor error"

representaciones para el caso de curvatura positiva hemos utilizado un "espacio circundante" —el espacio euclidiano en el que se encuentra el globo, o el espacio tridimensional en el que se muestra el espacio tiempo, fig. VII.11—, éste, el espacio circundante, no tiene una realidad física. El espacio en el interior o el exterior del globo está allí sólo para ayudarnos a visualizar su superficie. Es *únicamente* la superficie del globo la que representa el espacio físico del Universo. Vemos entonces que no hay punto central del que emane la materia del *big bang*. El "punto" en el centro del globo no es parte del universo, sino una ayuda para nuestra visualización del modelo, y el material que explotó está distribuido de modo uniforme por *todo* el espacio del universo.

La situación es la misma (aunque quizá un poco más difícil de visualizar) para los otros dos modelos estándar. El material jamás estuvo concentrado en un punto del espacio: desde el mismo comienzo llenaba uniformemente la *totalidad* del espacio.

Esta imagen está subyacente en la teoría del *big bang caliente* conocida como el *modelo estándar:* el Universo, instantes después de su creación, estaba en un estado térmico extremadamente caliente: la *bola de fuego primordial.*

Se han realizado cálculos detallados acerca de la naturaleza y proporciones de los constituyentes iniciales, y de cómo cambiaron estos constituyentes a medida que la bola de fuego (que era el Universo entero) se expandía y enfriaba. Puede parecer extraño que se puedan realizar cálculos fiables para describir un estado del Universo tan diferente del actual, pero la física sobre la que se basan estos cálculos no se pone en duda, siempre que no nos preguntemos qué sucedió antes de 10^{-4} segundos tras la creación. Desde ese instante: una diezmilésima de segundo después de la creación hasta unos tres minutos más tarde, se ha calculado el comportamiento con gran detalle *(cfr.* Weinberg, 1977) y, curiosamente, las teorías físicas —derivadas de un conocimiento experimental de un Universo ahora en un estado muy diferente— resultan adecuadas para ello.[7] Estos cálculos implican que —distribuidos de manera uniforme— por todo el Universo, debe haber muchos fotones (es decir, luz), electrones y protones (los dos constituyentes del hidrógeno), algunas partículas α (los núcleos de helio), un número aún menor de deuterones (los núcleos del deuterio, un isótopo pesado del hidrógeno), y trazas de otros tipos de núcleos, con quizá también un gran número de partículas "invisibles", tales como los neutrinos, que apenas dejarían sentir su presencia. Los constituyentes *materiales* (principalmente protones y electrones) se combinarían para producir el gas (principalmente hidrógeno) a partir del cual se formaron las estrellas... unos 10^8 años después del *big bang.*

Las estrellas no se formaron inmediatamente. Tras la expansión y enfriamiento del gas fueron necesarias concentraciones de éste en ciertas regiones para que los efectos gravitatorios locales pudieran superar la expansión global. ¿Cómo se formaron las galaxias y qué irregularidades tuvieron que estar presentes para que fuera posible su formación? No quiero entrar en esta discusión. Aceptemos, empero, que debió estar presente algún tipo de irregularidad en la distribución inicial del gas, y que el tipo correcto de agrupamiento gravitatorio se inició de tal

[7] Las bases experimentales para esta confianza proceden de dos tipos de datos: en primer lugar, el comportamiento de las partículas, cuando chocan entre sí velozmente para rebotar, fragmentarse y crear nuevas partículas, se conoce por medio de los aceleradores de partículas de alta energía construidos en varios lugares de la Tierra, y mediante el comportamiento de las partículas de los rayos cósmicos que inciden en la Tierra procedentes del espacio exterior. En segundo lugar, se sabe que los parámetros que gobiernan el modo de interacción de las partículas no han variado ni siquiera una parte en 10^6, en 10^{10} años *(cfr.* Barrow, 1988), por lo cual se considera altamente probable que no hayan cambiado de manera significativa (y probablemente nada en absoluto) desde el tiempo de la bola de *fuego* primordial.

modo que las galaxias pudieron formarse con sus cientos de miles de millones de estrellas constituyentes.

Hemos descubierto de dónde procede el gas difuso. Procedía de la misma bola de fuego que constituyó el *big bang*. El hecho de que este gas se haya distribuido de manera muy uniforme por todo el espacio es lo que nos ha dado la segunda ley, en la forma detallada en que ésta ha llegado a nosotros, después de que el procedimiento de elevación de la entropía de las masas gravitacionales fue disponible para nosotros. ¿Hasta qué punto *está* uniformemente distribuido el material que constituye el Universo? Las estrellas están agrupadas en galaxias, las galaxias también están agrupadas en cúmulos de galaxias y los cúmulos lo están en los llamados supercúmulos. Hay incluso alguna evidencia de que esos supercúmulos se reúnen en unos agolpamientos todavía mayores que son conocidos como complejos de supercúmulos. Y, no obstante, toda esa irregularidad es una minucia en comparación con la impresionante uniformidad de la estructura general del Universo. Cuanto más atrás — en el tiempo— ha sido posible mirar y cuanto mayor es la porción del universo explorado, más uniforme parece. La radiación de fondo de cuerpo negro proporciona la evidencia más impresionante a este respecto. Nos dice, en concreto, que cuando el Universo tenía apenas un millón de años, y sobre un dominio que se extendía a unos 10^{23} kilómetros (a una distancia de nosotros que abarcaría unas 10^{10} galaxias), el Universo y su contenido material eran uniformes hasta una parte en cien mil (*cfr.* Davies, 1987). El Universo, pese a sus orígenes violentos, era realmente muy uniforme ya desde sus etapas primitivas. Por consiguiente fue la bola de fuego inicial la que dispersó tan uniformemente este gas por el espacio.

¿EXPLICA EL BIG BANG LA SEGUNDA LEY?

¿Ha concluido nuestra búsqueda? ¿Está "explicado" el enigma de que la entropía de nuestro Universo empezara tan baja — el hecho que nos ha dado la segunda ley de la termodinámica — sólo por la circunstancia de que el Universo comenzó en un *big bang?* Hay algo paradójico en esta idea. No puede ser la verdadera respuesta. Recordemos que la bola de fuego primordial era un estado *térmico,* un gas caliente en equilibrio térmico en expansión. Recordemos también que el término "equilibrio térmico" designa el estado de *máxima entropía.* (Así es como designábamos al estado de máxima entropía de un gas en una caja.) Sin embargo, la segunda ley exige que en su estado inicial la entropía de nuestro Universo esté en alguna especie de *mínimo,* no en un máximo. ¿Qué es lo que anda mal?

Una respuesta "fácil" sería que, en efecto, la bola de fuego estaba en equilibrio térmico en el inicio, cuando el Universo era muy pequeño. Representaba el estado de entropía máxima permitido para un Universo de *tal* tamaño, pero la entropía entonces habría sido minúscula también, sobre todo en comparación con la que es posible para un Universo del tamaño del que hoy encontramos. A medida que el Universo se expandía, la entropía máxima permitida aumentó junto con el Universo, pero la entropía real del Universo quedó muy por debajo de ese máximo. La segunda ley aparece debido a que la entropía siempre trata de alcanzar este tamaño.

Sin embargo, un pequeño examen nos dice que ésta no puede ser la explicación correcta. Si lo fuera, entonces el argumento se aplicaría otra vez en la dirección *inversa* del tiempo, en el caso de un modelo de Universo (cerrado espacialmente) que colapse finalmente en un *big crunch.* Cuando el Universo alcanzara finalmente un tamaño minúsculo, comenzaría de nuevo a haber un

techo para los posibles valores de la entropía. La misma limitación que servía para darnos una baja entropía en las etapas primitivas del Universo en expansión se aplicaría de nuevo en las etapas finales del Universo en contracción. Era un límite para la baja entropía "en el principio del tiempo" el que nos daba la segunda ley, según la cual la entropía crece con el tiempo. Si este mismo límite para la baja entropía tuviera que aplicarse al "final del tiempo", entraríamos en conflicto con la segunda ley de la termodinámica. Por supuesto, también pudiera ser que nuestro Universo *real* no vaya a colapsar nunca más: Tal vez estamos viviendo en un Universo con curvatura espacial globalmente nula (caso euclidiano) o con curvatura negativa (caso de Lobachevsky), quizá *estamos* viviendo en un Universo (con curvatura positiva) que colapsa, pero ese colapso ocurrirá en un tiempo tan remoto que ninguna violación de la segunda ley sería discernible en nuestra época y la segunda ley, como hoy la entendemos, sería violada, además, desde esta perspectiva, llegaría el momento en que el incremento de la entropía *global* del Universo se detendría y finalmente iría disminuyendo hasta un valor pequeñísimo.

Existen buenas razones para dudar de que pudiera haber tal inversión de la entropía en un Universo colapsante. Algunas de las más poderosas de estas razones tienen que ver con esos misteriosos objetos conocidos como *agujeros negros.* En un agujero negro tenemos un microcosmos colapsante, de modo que si la entropía se invirtiera en un Universo colapsante entonces deberían darse también groseras violaciones de la segunda ley en las proximidades de un agujero negro. Sin embargo, existen todas las razones para creer que, en los agujeros negros, la segunda ley sigue firmemente en vigor. La teoría de los agujeros negros proporcionará un dato de partida vital para nuestro examen de la entropía.

AGUJEROS NEGROS

Consideremos primero lo que nos dice la teoría sobre el destino final de nuestro Sol. El Sol lleva existiendo unos cinco mil millones de años. Dentro de otros 5 o 6 mil millones de años empezará a expandirse en tamaño, hinchándose inexorablemente hasta que su superficie alcance la órbita de la Tierra. Entonces se habrá convertido en un tipo de estrella conocido como una *gigante roja.* Muchas gigantes rojas se pueden observar en otros lugares del cielo, siendo dos de las más conocidas Aldebarán en Tauro y Betelgeuse en Orión. Mientras su superficie esté expandiéndose, en su mismo núcleo habrá una pequeña concentración de materia excepcionalmente densa y en crecimiento continuo. Este núcleo denso tendrá la naturaleza de una estrella *enana blanca* (fig. VII 12). Las estrellas enanas blancas, propiamente dichas, son auténticas estrellas cuyo material está concentrado a una densidad tan alta que una bola de ping pong llena de este material pesaría varios cientos de toneladas. Estas estrellas se observan en el cielo en número considerable: quizá un diez por ciento de las estrellas de nuestra Vía Láctea sean enanas blancas. La enana blanca más famosa es la compañera de Sirio, cuya alarmante alta densidad supuso un gran enigma observacional para los astrónomos de principios de siglo. Con el paso del tiempo, sin embargo, esta misma estrella proporcionó una maravillosa confirmación de una teoría física (debida originalmente a R. H. Fowler, alrededor de 1926) según la cual algunas estrellas de tan alta densidad podrían mantenerse así por la "presión de degeneración electrónica", o sea que el principio mecánico-cuántico de exclusión de Pauli, aplicado a los electrones" impediría que la estrella colapsara gravitatoriamente hacia adentro.

Toda gigante roja tendrá una enana blanca en su núcleo central, y este núcleo absorberá continuamente material del cuerpo principal dé la estrella. Finalmente, la gigante roja habrá de ser consumida por este núcleo "parásito" y todo lo que quedará será una enana blanca —de un tamaño similar al de la Tierra.

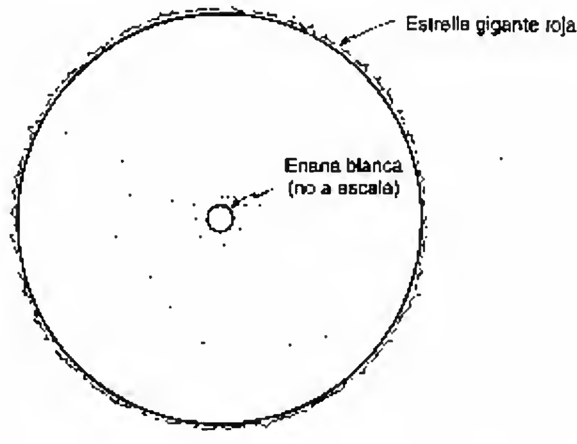


FIGURA VII 12. *Una gigante roja con una enana blanca en él núcleo.*

Se espera que nuestro Sol existirá como gigante roja durante "sólo" unos miles de millones de años, después de los cuales, en su última encarnación visible —como un rescoldo de enana blanca que se enfría* y agoniza lentamente—, persistirá durante unos pocos miles de millones de años más hasta llegar a una oscuridad total como una invisible *enana negra.*

Pero no todas las estrellas compartirán el destino del Sol. Para algunas su sino es mucho más violento y su futuro está decidido por lo que se conoce como el *límite de Chandrasekhar:* el máximo valor posible para la masa de una estrella enana blanca. Según un cálculo realizado en 1929 por Subrahmanyan Chandrasekhar, las enanas blancas no pueden existir si sus masas son, aproximadamente, de más de una vez y media la masa de nuestro Sol. (Cuando hizo este cálculo, viajando en barco desde la India a Inglaterra, él era un joven indio que se preparaba para ser investigador.) El cálculo fue repetido independientemente por el ruso Lev Landau hacia 1930. El valor moderno algo refinado para el límite de Chandrasekhar, es de aproximadamente

$$1.4\, M_{\odot},$$

donde $M_{\odot}$ es la masa del Sol, es decir, $M_{\odot}$ = una *masa solar.*

Nótese que el límite de Chandrasekhar no es mucho mayor que la masa del Sol, mientras que se conocen muchas estrellas ordinarias cuya masa es considerablemente mayor. ¿Cuál sería el destino final de una estrella de masa $2\, M_{\odot}$, por ejemplo?

De nuevo, según la teoría establecida, la estrella se hincharía hasta hacerse una gigante roja y su núcleo de tipo enana blanca adquiriría lentamente una masa igual que antes. Sin embargo, en algún momento crítico el núcleo alcanzará el límite de Chandrasekhar y el principio de exclusión de Pauli será insuficiente para mantenerla contra las enormes presiones inducidas por la

* De hecho, en sus etapas finales la enana brillará tenuemente como una estrella roja, aunque lo que se conoce como "enana roja" es una estrella de un carácter muy diferente.

gravitación.[8] En este punto, o próximo a él, el núcleo se colapsará catastróficamente hacia adentro y las temperaturas y presiones se incrementarán enormemente, en medio de violentas reacciones nucleares y una enorme cantidad de energía que se liberará del núcleo en forma de neutrinos. Éstos calentarán las regiones más externas de la estrella, las cuales habían estado colapsando hacia adentro, y se producirá una extraordinaria explosión. La estrella se transformará en una supernova.

¿Qué sucede entonces con el núcleo que aún sigue colapsado? La teoría dice que alcanza densidades todavía más gigantescas que las densidades alarmantemente altas que se daban en el interior de una enana blanca. El núcleo puede estabilizarse como una *estrella de neutrones* en la que es la *presión de degeneración neutrónica* —esto es, el principio de Pauli aplicado a los neutrones— la que mantiene a la estrella. La densidad será tal, que la misma bola de ping pong conteniendo material de estrellas de neutrones pesaría tanto como el asteroide Hermes (o quizá como Deimos, uno de los satélites de Marte). Este es el tipo de densidad que se encuentra en el interior de los propios núcleos atómicos. (Una estrella de neutrones es como un núcleo atómico gigantesco, quizá de unos diez kilómetros de radio que, sin embargo, es extremadamente pequeño para los niveles estelares.)

Pero ahora existe un *nuevo* límite, análogo al de Chandrasekhar (conocido como el límite de Landau-Oppenheimer-Volkov), cuyo valor moderno (revisado) es aproximadamente

$$2.5\,M_{\Theta}$$

por encima del cual no puede mantenerse una estrella de neutrones.

¿Qué le sucede a este núcleo colapsante si la masa de la estrella original es tan grande que incluso excede *este* límite?

Por ejemplo, se conocen muchas estrellas con masas comprendidas entre 10 M_{Θ} y 100 M_{Θ}. Parecería muy poco probable que arrojaran invariablemente tanta masa que el núcleo resultante quedara por debajo del límite para una estrella de neutrones. En lugar de ello, lo que se espera es que resulte un *agujero negro.*

¿Y qué es un agujero negro? Es una región del espacio —o del espacio-tiempo— en la que el campo gravitatorio es tan intenso que ni siquiera la luz puede escapar de ella. Recordemos que una consecuencia de los principios de la relatividad es que la velocidad de la luz es una velocidad límite: ningún objeto material o señal puede superar la velocidad local de la luz. En consecuencia, si la luz no puede escapar de un agujero negro, ¡*nada* puede escapar!

Quizá el lector esté familiarizado con el concepto de *velocidad de escape.* Esta es la velocidad que debe alcanzar un objeto para escapar de algún cuerpo con masa. Supongamos que este cuerpo fuera la Tierra: la velocidad de escape sería de aproximadamente 40000 kilómetros por hora, o algo más de 11 kilómetros por segundo. Una piedra que se lanzara desde la superficie de la Tierra (y que se alejara del suelo en cualquier dirección) con una velocidad que superara este valor, escaparía por completo de la Tierra (bajo el supuesto de que podemos despreciar los efectos de la resistencia del aire). Si la arrojamos con una velocidad menor que ésta, entonces caerá de nuevo en la superficie de la Tierra. (No es cierto, por lo tanto, que "todo lo que sube

[8] El principio de Pauli no prohíbe que los electrones estén en el mismo "lugar" sino que dos de ellos estén en el mismo "estado", que incluye también el modo en que los electrones se desplazan y giran. El argumento real es un poco delicado y fue objeto de mucha controversia, en particular por Eddington.

debe bajar". Un objeto retorna sólo si es arrojado con una velocidad *menor* que la velocidad de escape.) Para Júpiter, la velocidad de escape es de 220 000 kilómetros por hora; y para el Sol es de 2 200 000 kilómetros por hora.

Imaginemos que la masa del Sol estuviera concentrada en una esfera de sólo un cuarto de su radio actual, lo que supondría una velocidad de escape del doble. Si el Sol estuviera más concentrado, digamos en una esfera de *una centésima* parte de su radio actual, entonces la velocidad sería *diez veces* mayor. Podemos imaginar que para un cuerpo concentrado y de suficiente masa, la velocidad de escape superaría incluso a la velocidad de la luz. Cuando esto sucede, tenemos un agujero negro.[9]

La fig- VII. 13 representa un diagrama espacio-temporal que muestra el colapso de un cuerpo para formar un agujero negro (donde estoy suponiendo que el colapso tiene lugar de un modo que mantiene razonablemente cerca la simetría esférica, y donde he suprimido una de las dimensiones espaciales). Se han dibujado los conos de luz. Como lo vimos al examinar la relatividad general en el capítulo V, estos conos indican los límites absolutos para el movimiento de un objeto material o señal. Nótese que los conos empiezan a inclinarse hacia adentro y que su inclinación se hace mayor cuanto más próximos están al centro.

Hay una distancia crítica a partir del centro, llamada *radio de Schwarzschild,* en la que los límites exteriores de los conos se hacen *verticales* (en el diagrama). A esta distancia, lo más que puede hacer la luz (que debe seguir a lo largo de los conos de luz) es mantenerse sobre el cuerpo colapsado: toda la velocidad hacia afuera que la luz pueda reunir es apenas suficiente para contrarrestar la enorme atracción gravitatoria.

La superficie en el espacio-tiempo que describe, en el radio de Schwarzschild, esta luz que se mantiene sobre el cuerpo (es decir, la historia entera de la luz) se conoce como el *horizonte de sucesos (absoluto)* del agujero negro. Cualquier cosa que se encuentre dentro del horizonte de sucesos no puede escapar, ni siquiera comunicarse con el mundo exterior. Esto puede verse a partir de la inclinación de los conos y del hecho fundamental de que todos los movimientos y señales están obligados a propagarse adentro o a lo largo de ellos. Para un agujero negro formado por el colapso de una estrella de unas pocas masas solares el radio del horizonte será de unos cuantos kilómetros. Se cree que puede haber agujeros negros mucho mayores en los centros de las galaxias. Nuestra propia Vía Láctea podría contener un agujero negro de aproximadamente un millón de masas solares, y el radio del agujero sería entonces de algunos millones de kilómetros.

El cuerpo material real que colapsa para formar el agujero negro acabará en su totalidad dentro del horizonte, y por ello no puede comunicarse con el exterior. Consideraremos después el destino probable de ese

[9] Este razonamiento fue desarrollado en 1784 por el astrónomo inglés John Michell, y después por Laplace, de manera independiente. Ellos llegaron a la conclusión de que los cuerpos del Universo con más masa y más concentrados podrían ser totalmente invisibles —como agujeros negros—, pero debido a que sus argumentos (ciertamente proféticos) se basaban en la teoría newtoniana, sus conclusiones son por lo menos discutibles. Un tratamiento adecuado a partir de la relatividad general fue dado de manera independiente por John Robert Oppenheimer y Hartland Snyder en 1939.

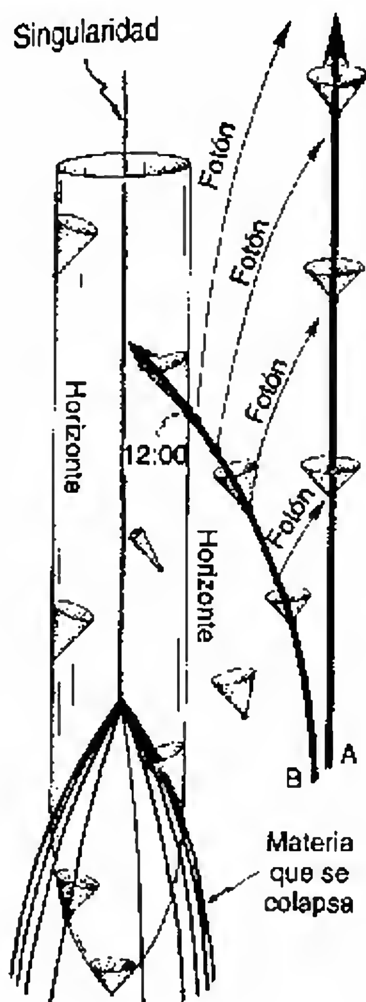


FIGURA VII. 13. *Diagrama espacio-temporal que representa el colapso de un agujero negro. El radio de Schwarzschild figura marcado como "Horizonte".*

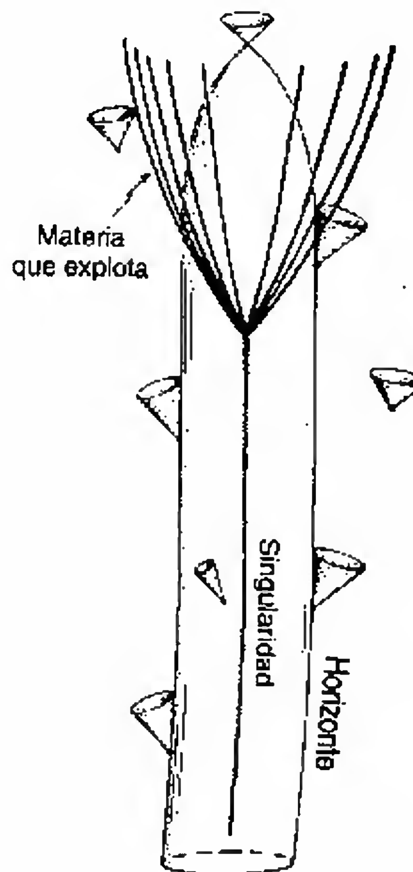


FIGURA VII 14. *Hipotética configuración espacio-temporal: un agujero blanco que finalmente explota en materia; el inverso en el tiempo del diagrama de la fig. VII 13.*

cuerpo. Por el momento, lo que nos interesa es solamente la geometría espacio-temporal creada por su colapso, una geometría espacio-temporal con curiosas y profundas implicaciones.

Imaginemos un valiente (¿o temerario?) astronauta B que decide viajar al interior del agujero negro, mientras su compañero A, más prudente, permanece a salvo, alejado. Supongamos que A se esfuerza por mantenerse a la vista de B el mayor tiempo posible. ¿Qué ve A? Puede pensarse, a partir de la fig. VII 13, que la porción de la historia de B (esto es, la línea de universo de B) que está en el *interior* del horizonte nunca será vista por A, mientras que la porción *afuera* del horizonte se hará finalmente visible a A, aunque los momentos inmediatamente anteriores a que B se adentre, a través del horizonte, serán vistos por A sólo tras periodos de espera cada vez mayores.

Supongamos que B cruza el horizonte cuando su propio reloj marca las 12 en punto. El astronauta A nunca podrá ser testigo de este acontecer, pero las lecturas del reloj 11:30, 11:45, 11:52, 11:56, 11:58, 11:59, 11:59$_{1/2}$, 11:59$_{3/4}$, 11:59$_{7/8}$, etc., serán vistas sucesivamente por A (a intervalos aproximadamente iguales, según el punto de vista de A).

En principio, B permanecerá siempre visible para A, y parecerá estar inmóvil justo siempre en el borde del horizonte, con su reloj acercándose con lentitud creciente a la hora fatídica de las 12:00, aunque sin alcanzarla. Pero, en realidad, la imagen de B que es advertida por A se haría rápidamente demasiado tenue para ser discernible. Esto se debe a que la luz de la pequeñísima porción de la línea de universo de B que queda fuera del horizonte es toda la que percibirá A durante el resto del tiempo. De hecho, B habrá desaparecido de la vista de A, lo mismo que todo el cuerpo colapsante original. Todo lo que A pueda ver será, precisamente, un agujero negro.

¿Qué le sucede al pobre B? ¿Cuál será *su* experiencia? Debe señalarse, en primer lugar, que no ocurrirá nada especial para B en el momento que cruce el horizonte. El podrá mirar su reloj y verá que los minutos pasan regularmente: 11:57, 11:58, 11:59, 12:00, 12:01, 12:02, 12:03,... Nada resulta especial en torno a las 12:00. Puede volverse para mirar a A y encontrará que éste permanece continuamente a la vista durante todo el tiempo, y puede mirar el propio reloj de A, que a B le parecerá que avanza de una forma ordenada y uniforme. A menos que B haya *calculado* que debe haber cruzado el horizonte, no tendrá medio de saberlo.[10] El horizonte ha sido extremadamente "perverso": una vez cruzado, ya no hay escape para B. Su Universo local colapsará sobre él, y encontrará rápidamente su *big crunch* particular.

O quizá no sea tan privado. Toda la materia del cuerpo colapsado, que formó el agujero negro en primer lugar, estará compartiendo, en cierto sentido, el "mismo" *crunch* con él. De hecho, si el Universo *afuera* del agujero está cerrado espacialmente, de modo que la materia exterior acabará también por ser engullida en un *big crunch* que englobe todo, es de esperar que dicho *crunch* sea el mismo que el particular de B.[*]

A pesar de su poco agradable destino, no esperarnos que la física local que B experimenta hasta ese punto pudiera estar reñida con la física que hemos llegado a conocer y comprender. En particular no esperamos que él experimente violaciones locales de la segunda ley de la termodinámica, ni mucho menos una completa inversión del comportamiento creciente de la entropía. La segunda ley seguirá siendo tan válida en el interior del agujero negro como lo es en cualquier otra parte. La entropía en la vecindad de B seguirá en incremento hasta el instante mismo de su *crunch* final.

Para entender cómo la entropía en un *big crunch* (ya sea individual o generalizado) puede ser enormemente alta, mientras que la entropía en el *big bang* tuvo que haber sido mucho más baja, necesitaremos ahondar todavía más en la geometría espacio-temporal de un agujero negro. Antes de hacerlo, el lector echará también una ojeada a la fig. VII 14 que muestra el hipotético inverso temporal de un agujero negro: un *agujero blanco*. Los agujeros blancos probablemente *no* existan en la naturaleza, pero su posibilidad teórica tiene importancia para todos.

---

[10] La localización *exacta* del horizonte, en el caso de un agujero negro general no estacionario, no puede determinarse con medidas directas. Depende en parte de un conocimiento de todo el material que caerá en el agujero negro... en su *futuro*.

[*] Al hacer esta afirmación estoy aceptando dos hipótesis. La primera es que la posible desaparición final del agujero negro — debida a su (extraordinariamente lenta) "evaporación" por radiación de Hawking que consideraremos más adelante — quedaría impedida por el colapso del Universo; la segunda es una célebre hipótesis conocida como censura cósmica .

LA ESTRUCTURA DE LAS SINGULARIDADES DEL ESPACIO-TIEMPO

Recordemos del capítulo V que la curvatura del espacio-tiempo se manifiesta como un *efecto de marea.* Una superficie esférica, constituida por partículas en caída libre en el campo gravitatorio de algún cuerpo grande, será estirada en una dirección (a lo largo de la línea dirigida hacia el cuerpo) y aplastada en direcciones perpendiculares a aquélla. Esta distorsión aumenta a medida que se acerca al cuerpo gravitatorio (fig. VII 15), variando de forma inversamente proporcional al cubo de la distancia.

Un efecto semejante de marea en aumento será sentido por el astronauta B a medida que cae hacia el agujero negro. Para un agujero negro de unas pocas masas solares este efecto de marea sería enorme, demasiado para que el astronauta pudiera sobrevivir a cualquier aproximación al agujero, y mucho menos al cruce del horizonte. Para agujeros más grandes, el tamaño del efecto de marea en el horizonte sería más pequeño. Para el agujero negro de millones de masas solares que muchos astrónomos creen que puede haber en el centro de la Vía *Láctea,* el efecto de marea sería suficientemente pequeño para que el astronauta cruzara el horizonte, aunque lo bastante grande para que se sintiera incómodo.

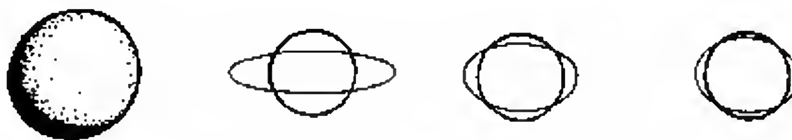


FIGURA VII 15. *El efecto de marea debido a un cuerpo esférico gravitante aumenta a medida que el cuerpo se acerca, en proporción al inverso del cubo de la distancia al centro del cuerpo.*

Sin embargo, este efecto de marea no se mantiene pequeño durante toda la caída del astronauta, sino que crece rápidamente a infinito en cuestión de unos cuantos segundos.

No sólo el desafortunado astronauta sería despedazado por esta fuerza de marea rápidamente creciente sino que también lo serían, en rápida sucesión, las moléculas de las que estaba compuesto, sus átomos constituyentes, sus núcleos e incluso, todas las partículas subatómicas Así es como la gran explosión lleva a cabo su último estrago.

No sólo toda la materia se destruye de este modo, sino que incluso el propio espacio-tiempo encuentra así su final. Semejante catástrofe es conocida como una *singularidad del espacio-tiempo.* El lector puede preguntarse cómo sabemos que deben ocurrir tales singularidades, y en qué circunstancias la materia y el espacio-tiempo sufren este destino. Son consecuencias de las ecuaciones clásicas de la relatividad general, en cualquier circunstancia en que se forme un agujero negro.

El modelo original de agujero negro de Oppenheimer y Snyder (1939) mostraba un comportamiento de este tipo; sin embargo, los astrofísicos conservan la esperanza de que este comportamiento singular sea un artificio de las simetrías espaciales que tenían que suponerse en el modelo. Quizá en situaciones realistas (asimétricas) la materia colapsante se enrosque de alguna forma complicada y luego escape de nuevo hacia afuera.

Tales esperanzas se fueron por tierra cuando se pudo disponer de los argumentos matemáticos, de tipo más general, que proporcionan los llamados *teoremas de singularidad (cfr.* Penrose,

1965; Hawking y Penrose, 1970). Estos teoremas establecieron, dentro de la teoría clásica de la relatividad con fuentes materiales razonables, que las singularidades en el espacio-tiempo son *inevitables* en situaciones de colapso gravitatorio.

Análogamente, si utilizamos la dirección inversa del tiempo, llegamos a una correspondiente singularidad *inicial* en el espacio-tiempo, que representa al *big bang,* en cualquier Universo en (apropiada) expansión.

Aquí, en lugar de representar la *destrucción* final de toda la materia del espacio-tiempo, la singularidad representa la *creación* de espacio-tiempo y materia.

Podría parecer que hay una simetría temporal exacta entre estos dos tipos de singularidad: el tipo *inicial,* en el que se crean el espacio-tiempo y la materia, y el tipo *final,* en el que se destruyen. Hay, en efecto, una importante similitud entre estas dos situaciones, pero *no* son exactamente inversas en el tiempo. Es importante que comprendamos las diferencias geométricas, porque ellas contienen la clave del origen de la segunda ley de la termodinámica.

Volvamos ahora a las experiencias de nuestro autosacrificado astronauta, quien encuentra que las fuerzas de marea crecen rápidamente hacia el infinito. Puesto que está viajando en un espacio vacío, él experimenta los efectos de conservación de volumen, aunque con *distorsión,* que son debidos al tipo de tensor de curvatura espacio-temporal que he llamado WEYL (véase capítulo V). La parte restante del tensor de curvatura espacio-temporal, la parte que representa una compresión global y a la que llamamos Ricci, es nula en el espacio vacío. Pudiera ser que, en efecto, B encontrara materia en alguna etapa, pero incluso si así fuera (y él mismo está constituido, después de todo, de materia) seguiríamos descubriendo que la medida de WEYL es mucho *mayor* que la de Ricci.

Esperamos encontrar que la curvatura cerca de una singularidad *final* esté dominada completamente por el tensor WEYL. Este tensor tiende a *infinito,* en general:

$$\text{WEYL} \rightarrow \infty$$

(aunque podría hacerlo perfectamente de manera oscilatoria).

Esta parece ser la situación *genérica* con una singularidad en el espacio-tiempo.[11] Tal comportamiento está asociado con una singularidad de *alta entropía.* Sin embargo, la situación con el *big bang* parece ser diferente.

Los modelos estándar del *big bang* se derivan de los espacio-tiempos altamente simétricos de Friedmann-Robertson-Walker que consideramos anteriormente. El efecto de marea distorsionante que proporciona el tensor WEYL está totalmente *ausente* y, en su lugar hay una aceleración simétrica hacia adentro que actúa en cualquier superficie esférica de partículas de prueba (véase fig. V.26). Ésta es un efecto del tensor RICCI, más que del WEYL. En cualquier modelo FRW siempre se satisface la ecuación tensorial

$$\text{WEYL} = 0$$

Cada vez que nos aproximamos a la singularidad inicial, encontramos que es RICCI el que se hace infinito, en lugar de WEYL, de modo que es RICCI el que domina cerca de la singularidad inicial, por encima de WEYL Esto nos proporciona una singularidad de *baja entropía.*

[11] Véanse las discusiones de Belinskii, Khalatnikov y Lifshitz (1970) y Penrose (1979b).

Si examinamos la singularidad del *big crunch* en los modelos FRW *exactamente* recolapsantes, encontramos WEYL = 0 en el momento final del *crunch,* mientras que RICCI tiende a infinito. Sin embargo, ésta es una situación muy especial y *no* es lo que esperamos para un modelo realista en el que se tenga en cuenta el agrupamiento gravitatorio. A medida que *avanza* el tiempo, el material, originalmente en forma de gas difuso, se agrupará en constelaciones. Y a su debido tiempo, un gran número de esas estrellas se contraerán gravitatoriamente para derivar en enanas blancas, estrellas de neutrones y agujeros negros, y podrían darse incluso agujeros negros en los centros de las constelaciones o galaxias. El agrupamiento —particularmente en el caso de los agujeros negros— representa un enorme incremento de entropía (véase fig. VII 16).

Puede parecer enigmático, al principio, que los estados agrupados representen *alta* entropía y los más uniformes baja entropía, cuando recordamos que, en el caso del gas en una caja, los estados agrupados (como cuando éste estaba concentrado en una esquina) eran de *baja* entropía, mientras que en el estado uniforme de equilibrio térmico la entropía era *alta.*

Cuando consideramos la gravedad se produce la situación *inversa,* debido a la naturaleza fundamentalmente atractiva del campo gravitatorio. El agrupamiento se hace extremo conforme pasa el tiempo y, finalmente, se coagulan los agujeros negros y sus singularidades se unen en la muy compleja singularidad final del *big crunch.*

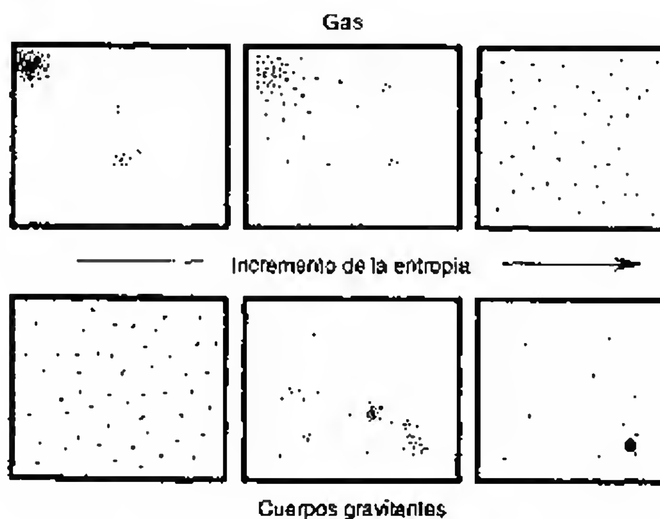


FIGURA VII. 16. *Para un gas ordinario, el incremento de la entropía tiende a hacer más uniforme la distribución. Para un sistema de cuerpos gravitantes sucede lo contrarío. La entropía alta se logra mediante agrupamiento gravitatorio —y la más alta, de todas, mediante el colapso de un agujero negro.*



FIGURA VII 17. *La historia completa de un universo cerrado que comienza en un* big bang *uniforme de baja entropía, con* WEYL = 0, *y termina en un* big crunch *de alta entropía —que representa la coagulación de muchos agujeros negros— con* WEYL $\rightarrow \infty$

La singularidad final no se parece al *big crunch* idealizado del modelo FRW recolapsante, con su ligadura WEYL = 0. A medida que se produce el agrupamiento, surge una tendencia del tensor Weyl a crecer cada vez más [12] y, en general, WEYL $\rightarrow \infty$ en cualquier singularidad final. Véase la fig. VII 17 para una imagen espacio-temporal que representa la historia de un Universo cerrado acorde con esta descripción.

Vemos ahora cómo es posible que un Universo recolapsado *no* necesite tener una pequeña entropía.

La *baja* entropía en el *big bang,* lo que nos dio la segunda ley, no es una mera consecuencia de la pequeñez del Universo en el momento del *big bang.* Si invirtiéramos en el tiempo la imagen del *big crunch* que obtuvimos antes, obtendríamos un *big bang* con una entropía enormemente *alta,* y no habría existido segunda ley. El Universo fue creado en un estado muy especial (de baja entropía), en el que se impuso algo parecido a la ligadura WEYL = 0 de los modelos FRW. Si no fuera por una ligadura

[12] Resulta tentador identificar la contribución gravitatoria a la entropía de un sistema con alguna medida de la curvatura total de Weyl, pero no se han encontrado medidas apropiadas. (En general tendría que tener algunas complejas propiedades no locales.) No necesitamos ahora esta medida de la entropía gravitatoria.



FIGURA VII 18. Si *se elimina la ligadura* WEYL = *O, tendremos también un* big bang *de alta entropía con* WEYL $\rightarrow \infty$. *Un universo de este tipo estaría surcado de agujeros blancos y no existiría la segunda ley de la termodinámica. Tal situación contradice la experiencia.*

de esta naturaleza, sería *mucho más probable* tener una situación en la que *ambas* singularidades (inicial y final) fueran del tipo WEYL $\rightarrow \infty$ de alta entropía (véase fig. VII.18). En un Universo *probable* así, no existiría, de hecho, la segunda ley de la termodinámica.

¿HASTA QUÉ PUNTO FUE ESPECIAL EL BIG BANG?

Tratemos de comprender hasta qué punto era restrictiva para el *big bang* una condición como WEYL = 0. Por simplicidad, supongamos que el Universo es cerrado. Para obtener un número redondo supondremos, además, que el número B de *bariones* —es decir el número de protones y neutrones juntos— en el Universo viene dado aproximadamente por

$$B = 10^{80}.$$

(No existe una *razón* especial para esta cifra, aparte del hecho de que para propósitos de observación B debe ser de *al menos ese* orden. Eddington afirmó en cierta ocasión haber calculado B exactamente, y obtuvo una cifra próxima al valor anterior. Nadie más cree ese cálculo particular, pero parece que se ha retenido el valor 10^{80}.)

Si se tomara *B mayor* que ese valor (y quizá en realidad tengamos $B = \infty$); las cifras serían todavía *más* sorprendentes que las extraordinarias cifras a que vamos a llegar en un minuto.

Tratemos de imaginar el espacio de fases del Universo *entero*. Cada punto en ese espacio de fases representa una diferente forma en que pudiera haber comenzado el Universo. Vamos a representar a Dios provisto de una "aguja" que será colocada en algún punto del espacio de fases (fig. VII. 19). Cada colocación diferente de la aguja proporciona un Universo diferente. Ahora bien, la precisión necesaria para la puntería del Creador depende de la entropía del Universo que se vaya a crear.

Sería relativamente "fácil" producir un Universo de alta entropía, puesto que para ello habría un gran volumen del espacio de fases donde colocar la aguja. Recordemos que la entropía es proporcional al logaritmo del volumen que interesa en el espacio de fases, pero para crear el Universo en un estado de baja entropía — de modo que realmente exista una segunda ley de la termodinámica — Dios debe apuntar a un volumen muchísimo más pequeño del espacio de fases. ¿Qué tan pequeña debería haber sido esa región, ya que el resultado es el Universo en que vivimos? Acudamos a una fórmula muy importante, debida a Jacob Bekenstein (1972) y Stephen Hawking (1975), que nos dice cuál debe ser la entropía de un *agujero negro:*

Consideremos un agujero negro y supongamos que el área de su horizonte es A. La fórmula de Bekenstein-Hawking para la entropía del agujero negro es

$$S_{bh} = \frac{A}{4} \times \left(\frac{kc^3}{G\hbar} \right)$$

donde *k* es la constante de Boltzmann, c es la velocidad de la luz, G es la constante gravitatoria de Newton, y $\hbar$ es la constante de Planck dividida por *2n.* La parte esencial de esta fórmula es el factor *A/4.* La parte entre paréntesis consta simplemente de las constantes físicas apropiadas.

En definitiva, la entropía del agujero negro es proporcional al área de su superficie. Para un agujero negro con simetría esférica, tal área resulta ser proporcional al cuadrado de la masa del agujero

$$A = m^2 \times 8\pi \left(\frac{G^2}{c^4} \right)$$

Si introducimos esto en la fórmula de Bekenstein-Hawking, hallaremos que la entropía de un agujero negro es proporcional al cuadrado de su masa:

$$S_{bh} = m^2 \times 2\pi \left(\frac{kG}{\hbar c} \right)$$

Así, la *entropía por unidad de masa* (S_{bh}/m) de un agujero negro es proporcional a su masa y, por lo tanto, es mayor para agujeros negros cada vez más grandes. En consecuencia, para una cantidad dada de masa (o equivalentemente, por la relación $E = mc^2$ de Einstein, para una cantidad dada de *energía)* la entropía mayor se alcanzará cuando todo el material colapse por completo un agujero negro. Más aún, dos agujeros negros ganan (enormemente) en entropía cuando se "engullen" mutuamente y dan lugar a un agujero negro unido. Los agujeros negros grandes, corno los que probablemente se encuentren en los centros de las galaxias, proporcionarán cantidades de entropía muchísimo mayores que los que encontramos en cualquier otra situación física.

Es necesario hacer una ligera puntualización en torno a la afirmación de que la mayor entropía se logra cuando toda la masa se concentra en un agujero negro. El análisis de Hawking de la termodinámica de los agujeros negros muestra que existirá también una *temperatura* no nula asociada a un agujero negro.

Una consecuencia de esto es que no toda la masa-energía puede estar contenida dentro del agujero negro en el estado de máxima entropía porque la máxima entropía se alcanza en equilibrio con un "baño térmico de radiación". La temperatura de esta radiación es ínfima para un agujero negro de tamaño razonable. Por ejemplo, para un agujero negro del tamaño de una masa como la de nuestro Sol, esta temperatura sería de unos 10^{-7} K, o sea algo más pequeña que la temperatura más baja que se ha medido en un laboratorio hasta la fecha, y muchísimo menor que la temperatura de 2.7 K del espacio intergaláctico. Para agujeros negros mayores, la temperatura de Hawking es todavía más pequeña.

La temperatura de Hawking sería importante para nosotros sólo en uno de estos dos casos: *i)* podrían existir en nuestro Universo agujeros negros muchísimo más pequeños, conocidos como *miniagujeros negros; ii)* el Universo no colapsa antes del *tiempo de evaporación de Hawking,* el tiempo que tardaría el agujero negro en evaporarse completamente. Con respecto a *i),* los miniagujeros negros podrían producirse únicamente en un *big bang* adecuadamente caótico. Tales miniagujeros no sólo no son numerosos en nuestro Universo real, sino que, de acuerdo con el punto de vista que estoy exponiendo aquí, deberían estar totalmente ausentes. Con respecto a *ii),* para un agujero negro de masa solar el tiempo de evaporación de Hawking sería de unas 10^{54} veces la edad actual del Universo, y para agujeros negros más grandes sería considerablemente más largo. Estos efectos no modifican la sustancia de los argumentos anteriores.

Para darnos una idea de la enormidad de la entropía de los agujeros negros consideremos lo que previamente pensábamos que suministraba la mayor contribución a la entropía del Universo: la radiación de fondo de cuerpo negro de 2.7 K. Los astrofísicos quedaron sorprendidos ante la enorme cantidad de entropía contenida en esa radiación, que está muy por encima de las cifras normales de entropía que encontramos en otros procesos *(v.g.,* en el Sol). La entropía de la radiación de fondo es del orden de 10^8 por cada barión (en donde estoy escogiendo unidades naturales, de modo que la constante de Boltzmann es la unidad). (Esto significa, de hecho, que

hay 10^8 fotones en la radiación de fondo por cada barión.) Así, con 10^{80} bariones en total tendremos una entropía de

$$10^{88}$$

para la entropía de la radiación de fondo en el Universo.

En realidad, si no fuera por los agujeros negros esta cifra representaría la entropía total del Universo, puesto que la entropía de la radiación de fondo supera con mucho la de cualesquiera otros procesos ordinarios: la entropía por barión en el Sol, por ejemplo, es del orden de la unidad, y por el contrario, para los niveles de los *agujeros negros* la entropía de la radiación de fondo es el "chocolate del loro". En efecto, la fórmula de Bekenstein-Hawking nos dice que la entropía por barión en un agujero negro de masa solar es del orden de 10^{20}, en unidades naturales, de modo que si el Universo hubiera consistido solamente en agujeros negros de esa masa, la cifra global habría sido mucho mayor que la dada más arriba, a saber

$$10^{100}.$$

El Universo no está constituido de esa forma, pero esta cifra empieza a decirnos qué pequeña debe de ser considerada la entropía de la radiación de fondo cuando los efectos implacables de la gravedad empiezan a tenerse en cuenta.

Seamos más realistas: en lugar de poblar completamente las galaxias con agujeros negros, consideremos que en lo fundamental constan de estrellas ordinarias —unas 10^{11} de ellas— y que cada galaxia tiene en su núcleo un agujero negro de un millón (esto es, 10^6) de masas solares, como podría ser razonable para nuestra propia Vía Láctea. Los cálculos muestran que la entropía por barión sería entonces incluso algo mayor que la cifra anterior, a saber 10^{21}, y que daría una entropía total, en unidades naturales, de

$$10^{101}.$$

Podemos prever que, después de un tiempo largo, una fracción importante de las masas de las galaxias estará incorporada en los agujeros negros de sus centros. Cuando esto suceda, la entropía por barión **será** 10^{31}, y dará un total monstruoso de

$$10^{111}.$$

Sin embargo, estamos considerando un Universo cerrado que colapsa, y no es irrazonable estimar la entropía del *crunch* final utilizando la fórmula de Bekenstein-Hawking como si todo el Universo hubiera formado un agujero negro.

Esto da una entropía por barión de 10^{43}, y el total —inimaginable—, para el *big crunch* sería

$$10^{123}.$$

Esta cifra nos dará una estimación del volumen total V del espacio de fases disponible al Creador, ya que representaría el logaritmo del volumen del (con mucho) mayor compartimento. Puesto que 10^{123} es el *logaritmo* del volumen, este volumen debe ser la exponencial de 10^{123}, es decir

$$V=10^{10^{123}}$$

en unidades naturales. (Algunos lectores perspicaces pueden pensar que debería haber utilizado la cifra $e^{10^{123}}$, pero para números de esta magnitud, la e y el 10 son prácticamente intercambiables).

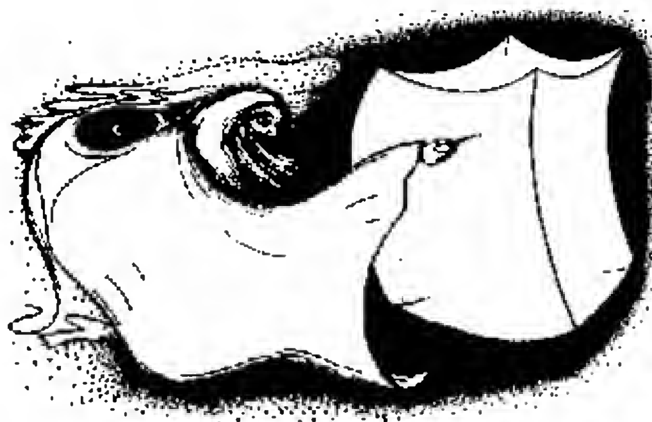


FIGURA VII 19. *Para producir un universo parecido al que habitamos, Dios tendría que haber apuntado a un volumen absurdamente minúsculo del espacio de fases de los universos posibles —aproximadamente* $1/10^{10^{123}}$ *del volumen total, para la situación considerada. (La aguja y el punto no están dibujados a escala.)*

¿Cuál era el volumen W del espacio de fases original que previo el Creador para proporcionar un Universo compatible con la segunda ley de la termodinámica y con el que ahora observamos? No importa mucho si tomamos el valor

$$W = 10^{10^{101}} \text{ o } \quad W = 10^{10^{88}}$$

dado por los agujeros negros galácticos o por la radiación de fondo, respectivamente, o una cifra mucho más pequeña (y, de hecho, más adecuada), que hubiera sido la cifra *real* en el *big bang*. En cualquier caso, la razón de V a W será, aproximadamente,

$$V/W = V=10^{10^{123}}$$

(Inténtelo: $10^{10^{123}} \div 10^{10^{101}} = 10^{(10^{123} - 10^{101})} = 10^{10^{123}}$ aproximadamente.)

Esto nos dice lo precisa que debía haber sido la puntería del Creador: una precisión "divina" de

$$\text{una parte en } 10^{10^{123}}$$

¡Una cifra extraordinaria! Ni siquiera podríamos *escribir el número* completo en la notación decimal ordinaria: sería un "1" seguido de 10^{123} "0"s. Incluso si escribiéramos un "0" en cada protón y en cada neutrón del Universo entero —y añadiéramos también todas las demás partículas—, todavía nos quedaríamos muy cortos. Se puede ver que la precisión necesaria para poner al Universo en su curso no es en modo alguno inferior a la extraordinaria precisión a la que ya nos habíamos acostumbrado en las ecuaciones dinámicas soberbias que gobiernan el comportamiento de las cosas (las de Newton, las de Maxwell, las de Einstein).

Pero, ¿por qué estaba el *big bang* tan exactamente organizado, mientras que el *big crunch (o* las singularidades de los agujeros negros) era totalmente caótico? Esta pregunta podríamos enunciarla en términos del comportamiento de la parte WEYL del tensor de curvatura espacio-temporal, en las singularidades del espacio-tiempo.

Lo que encontramos es una ligadura

$$\text{WEYL} = 0$$

(o algo muy parecido a esto) en las singularidades iniciales del espacio-tiempo —pero no en las singularidades finales—, y eso parece ser lo que confina la elección del Creador a esta región minúscula del espacio de fases.

Yo he llamado *hipótesis de curvatura de Weyl* a la hipótesis de que tal ligadura *se* aplica a cualquier singularidad inicial (pero no final) del espacio-tiempo. Parecería así que necesitamos entender por qué debería aplicarse esta hipótesis —temporalmente asimétrica— si tuviéramos que comprender de dónde procede la segunda ley.[13]

¿Cómo podemos obtener una comprensión mayor del origen de la segunda ley? Hemos llegado a un punto muerto. Necesitamos comprender por qué las *singularidades del espacio-tiempo* tienen las estructuras que parecen tener, pero estas singularidades del espacio-tiempo son regiones en las que nuestra comprensión de la física ha llegado al límite. El punto muerto que supone la existencia de singularidades en el espacio-tiempo se compara a veces con otro punto muerto: el que encontraron los físicos a principios de siglo, a propósito de la estabilidad de los átomos .

En cada caso, la teoría clásica bien establecida había llegado a la respuesta "infinito", y por consiguiente se había mostrado inadecuada para la tarea. El comportamiento singular del colapso electromagnético de los átomos era predicho por la teoría *cuántica,* y por ello debería ser la teoría cuántica la que diera una teoría finita en lugar de las singularidades "infinitas" clásicas en el espacio-tiempo para el colapso gravitatorio de las estrellas.

No se trataría, empero, de una teoría cuántica ordinaria. Debía ser una teoría cuántica de la propia estructura del espacio-tiempo. Tal teoría, si existiera, sería conocida como *"gravitación cuántica ".*

El hecho de que no exista una gravitación cuántica no se debe a falta de esfuerzo, capacidad o ingenio. Muchos cerebros de científicos de primera fila se han dedicado a la construcción de una teoría semejante, aunque sin éxito. Este es el punto muerto al que hemos llegado finalmente en nuestros intentos por comprender la direccionalidad y el flujo del tiempo.

El lector puede preguntarse qué es lo que hemos sacado en limpio de nuestro viaje. En nuestro intento por comprender por qué el tiempo fluye en una sola dirección, hemos tenido que viajar a sus mismos confine, donde las nociones mismas del espacio de disuelven. ¿Qué hemos aprendido de todo esto? Hemos aprendido que nuestras teorías no son todavía adecuadas para proporcionar respuestas. Pero, ¿nos ha servido en nuestro intento de comprender la mente? Pese a la falta de una teoría adecuada, hubo importantes lecciones. Ahora debemos encaminarnos de vuelta a casa. Nuestro viaje de regreso será más especulativo que el de ida.

---

[13] Existe un punto de vista muy popular en estos momentos, conocido como el "escenario inflacionario", que se propone explicar por qué el Universo es tan uniforme a gran escala. Según este punto de vista, el Universo sufrió una enorme expansión en sus primeras etapas, de un orden mucho mayor que la expansión "ordinaria" del modelo estándar. La idea es que cualquier irregularidad pudo ser reforzada por esa expansión; sin embargo, sin alguna ligadura inicial aún mayor, como la que ya proporciona la hipótesis de curvatura de Weyl, la inflación no puede funcionar: no introduce ningún ingrediente con asimetría temporal que pudiera explicar la diferencia entre las singularidades inicial y final. Además está basado en teorías físicas no confirmadas —las teorías GUT— cuyo estatus no pasa de "tentativas" en la terminología del capítulo V. Para un examen crítico de la "inflación" en el contexto de las ideas de este capítulo, véase Penrose, 1989b.

VIII. EN BUSCA DE LA GRAVITACIÓN CUÁNTICA

¿POR QUÉ LA GRAVITACIÓN CUÁNTICA?

¿QUÉ NUEVAS COSAS podemos aprender acerca de cerebros o mentes, aparte de lo que hemos visto en el último capítulo? Aunque podamos haber captado algunos de los principios físicos omnicomprensivos —que subyacen a la direccionalidad de nuestro "flujo del tiempo" percibido—, no hemos sacado ninguna idea sobre la cuestión de por qué percibimos que el tiempo fluye o, en realidad, por qué percibimos siquiera cualquier cosa.

En mi opinión, se necesitan ideas mucho más radicales. Hasta ahora mi presentación no ha sido especialmente radical, aunque he acentuado aspectos diferentes de los usuales.

Hemos tenido conocimiento de la segunda ley de la termodinámica y he intentado persuadir al lector de que en la raíz de esta ley —que nos presenta la naturaleza en la forma particular en que ha sido escogida— puede rastrearse hasta un estrecho vínculo geométrico en el *big bang* origen del Universo: *la hipótesis de curvatura de Weyl.* Algunos cosmólogos preferirían caracterizar esta condición inicial de forma algo diferente, pero tal restricción de la singularidad inicial es realmente necesaria; las *deducciones* que voy a obtener con la citada hipótesis serán mucho menos convencionales que ella misma. Necesitamos un cambio en el propio marco de la teoría cuántica. Este cambio jugará su papel cuando la mecánica cuántica se unifique apropiadamente con la relatividad general, es decir, en la tan deseada teoría de la *gravitación cuántica,*

La mayoría de los físicos no cree, hoy, que la teoría cuántica tenga que cambiar cuando se unifique con la relatividad general. Estos mismos físicos mantendrán mañana que, en la escala importante para nuestros cerebros, los efectos de *cualquier* gravitación cuántica deben ser completamente insignificantes. Dirán (muy razonablemente) que aunque tales efectos físicos podrían ser importantes en la absurdamente minúscula escala de la distancia conocida como *longitud de Planck*[*] —que es 10^{-35} m, unas 100 000 000 000 000 000 000 veces más pequeña que el tamaño de la más pequeña partícula subatómica—, estos efectos no tendrían ninguna importancia directa para fenómenos en escalas muchísimo más grandes de, pongamos por caso, hasta sólo 10^{-12}m, donde tienen lugar los procesos químicos o eléctricos que son importantes para la actividad cerebral.

De hecho, ni siquiera la gravitación *clásica* (esto es, no cuántica) tiene apenas importancia en esas actividades eléctricas y químicas. Si la gravitación clásica no tiene consecuencias, entonces ¿cómo se podría suponer la más mínima diferencia producto de cualquier minúscula "corrección cuántica" a la teoría clásica? Además, puesto que nunca se han observado *desviaciones* de la teoría cuántica, parecería todavía *menos* razonable el imaginar que cualquier supuesta minúscula desviación de la teoría cuántica estándar pudiera tener un papel que jugar en los fenómenos mentales.

Yo voy a argumentar de forma muy diferente. No estoy interesado tanto en los efectos que la mecánica cuántica pudiera tener en nuestra teoría de la estructura del espacio-tiempo (la

[*] Esta es la distancia $10^{-35}\, m = \sqrt{(hGc^{3}}\ )$ para la que las llamadas "fluctuaciones cuánticas" de la propia métrica del espacio-tiempo son tan grandes, que la idea normal de un continuo espacio-temporal suave deja de aplicarse. (Las fluctuaciones cuánticas son una consecuencia del principio de incertidumbre de Heisenberg.)

relatividad general de Einstein), como en el caso *inverso,* a saber: los efectos que la teoría del espacio-tiempo de Einstein pudiera tener sobre la propia estructura de la mecánica cuántica.

Recalcaré que es un punto de vista *no convencional* el que voy a desarrollar. No es convencional que la relatividad general pudiera tener alguna influencia sobre la estructura de la mecánica cuántica. Los físicos convencionales han sido muy renuentes a aceptar que la estructura estándar de la mecánica cuántica debería ser modificada de alguna forma. Aunque es cierto que la aplicación directa de las reglas de la teoría cuántica a la teoría de Einstein ha encontrado dificultades aparentemente insuperables, la reacción de quienes trabajan en este campo ha tendido a utilizar esto como una razón para modificar la teoría *de Einstein,* no la teoría cuántica.[1] Mi punto de vista es prácticamente el opuesto. Creo que los problemas internos de la teoría cuántica son fundamentales.

Recordemos la incompatibilidad entre los procedimientos básicos U y R de la mecánica cuántica (U obedece la completamente determinista *ecuación de Schrödinger* —llamada evolución *unitaria*—, y R era la probabilista *reducción del vector de estado* que debemos aplicar siempre que se estime que se ha hecho alguna "observación")- En mi opinión, esta incompatibilidad es algo que *no puede* resolverse mediante la simple adopción de una "interpretación" apropiada de la mecánica cuántica (aunque algo puede hacer), sino sólo mediante una teoría radicalmente nueva, en la cual ambos procedimientos, **U** y **R**, se verán como dos aproximaciones diferentes (y excelentes) a un procedimiento *único* más general y exacto.

Mi opinión, en consecuencia, es que incluso la maravillosamente exacta teoría de la mecánica cuántica tendrá que ser cambiada, y que los indicios sobre la naturaleza de este cambio tendrán que venir de la teoría de la relatividad general de Einstein. Iré todavía más lejos: será la nueva teoría de la *gravitación cuántica* la que contenga, como uno de sus ingredientes fundamentales, este supuesto proceso combinado **U/R**.

Desde un punto de vista *convencional,* por el contrario, cualquier implicación directa de la gravitación cuántica sería de una naturaleza más esotérica. He mencionado las expectativas de una alteración fundamental de la estructura del espacio-tiempo en la dimensión ridículamente pequeña de la longitud de Planck. Existe también la creencia (justificada, en mi opinión) de que la gravitación cuántica debería estar involucrada de modo fundamental en la determinación última de la naturaleza del "zoo" de las "partículas elementales" actualmente observadas. Por el momento no existe, por ejemplo, ninguna buena teoría que explique por qué las masas de las partículas deben ser las que son, cuando la "masa" es un concepto íntimamente ligado con el concepto de gravitación. (De hecho, la masa actúa unívocamente como la "fuente" de la gravitación.) Existen también buenas expectativas para que (según una idea desarrollada alrededor de 1955 por el físico sueco Oskar Klein) la teoría correcta de la gravitación cuántica sea útil para eliminar los infinitos que plagan la convencional teoría cuántica de campos .

La física es una unidad, y la *verdadera* teoría de la gravitación cuántica, cuando finalmente llegue, debe constituir ciertamente una parte profunda de nuestra comprensión detallada de las

[1] Estas modificaciones populares son: *i)* cambiar las ecuaciones reales de Einstein RICCI = ENERGÍA (vía "lagrangianos de alto orden"); *ii)* cambiar el número de dimensiones del espacio-tiempo de cuatro a algún número mayor (como en las llamadas "teorías de tipo Kaluza-Klein"); *iii)* introducir "supersimetría" (una idea tomada en préstamo del comportamiento cuántico de bosones y fermiones, combinada en un amplio esquema y aplicada, no de forma completamente lógica, a las coordenadas del espacio-tiempo); *iv)* teoría de cuerdas (un esquema radical muy popular en este momento en el que las "líneas de universo" se reemplazan por "historias de cuerdas" —combinada normalmente con las ideas de *ii)* y *iii).* Todas estas propuestas, por su popularidad y vigorosa presentación, entran firmemente en la categoría de TENTATIVAS en la terminología del capítulo V.

leyes universales de la naturaleza. Estamos, sin embargo, lejos de tal comprensión, y cualquier supuesta teoría de la gravitación cuántica debe permanecer alejada de los fenómenos que gobiernan el comportamiento del cerebro. *Especialmente* alejado de la actividad cerebral parecería estar ese papel (generalmente aceptado) de la gravitación cuántica que se necesita para resolver el punto muerto al que llegamos en el último capítulo: el problema de las *singularidades del espacio-tiempo* —las singularidades de la teoría clásica de Einstein que aparecen en el *big bang* y en los *agujeros negros*— y también en el *big crunch,* si nuestro Universo "decide" finalmente colapsar sobre sí mismo.

Sí, este papel podría *parecer* muy lejano. Argumentaré, de todos dos, que existe un hilo de conexión lógica, sutil pero importante. Trataremos de ver cuál es tal conexión.

¿QUÉ HAY DETRÁS DE LA HIPÓTESIS DE CURVATURA DE WEYL?

Como he señalado, incluso el punto de vista convencional nos dice que debería ser la gravitación cuántica la que viniera en ayuda de la teoría clásica de la relatividad general y resolver el misterio de las singularidades del espacio-tiempo. Así, la gravitación cuántica tiene que proporcionarnos alguna física coherente, en lugar de la absurda respuesta del "infinito" que nos da la teoría clásica. Coincido en esa opinión: éste es realmente un lugar en donde la gravitación cuántica debe dejar clara su huella. Sin embargo, los teóricos no parecen dispuestos a ponerse de acuerdo en el hecho de que la huella de la gravitación cuántica es descaradamente tiempo-asimétrica. En el *big bang —singularidad del pasado—* la gravitación cuántica tiene que decirnos que debe darse una condición similar a

$$\text{WEYL} = 0$$

en el momento que se haga significativo el hablar en términos de conceptos clásicos de geometría del espacio-tiempo. Por otro lado, en la singularidades en el interior de los agujeros negros, y en el (posible) *big crunch —singularidades futuras—* no existe tal restricción, y esperamos que el tensor de Weyl se haga infinito:

$$\text{WEYL} \rightarrow \infty$$

a medida que nos acercamos a la singularidad. En mi opinión, ésta es una indicación clara de que la teoría real que buscamos debe ser asimétrica respecto al tiempo:

nuestra buscada gravitación cuántica es una teoría tiempo-asimétrica.

Se advierte aquí al lector que esta conclusión, pese a su aparente necesidad obvia según el modo en que estoy presentando las cosas, ¡*no es* un saber aceptado! La mayoría de los que trabajan en este campo parecen muy reacios a aceptar esta idea. La razón parece ser el hecho de que no hay una manera clara mediante la que los procedimientos convencionales y bien establecidos de cuantización (hasta donde puedan llegar) pudieran producir una teoría cuantizada con asimetría temporal,[2] cuando la propia teoría clásica a la que se aplican estos procedimientos (la relatividad general estándar o una de sus modificaciones más conocidas) tiene simetría temporal. En consecuencia (cuando ellos consideran estos temas, ¡lo que no es muy frecuente!), estos

[2] Aun cuando la simetría de una teoría clásica no siempre es conservada por los procedimientos de cuantización *(cfr.* Treiman, 1985; Ashtekar *et al,* 1989), lo que se requiere aquí es una violación de las *cuatro* simetrías comúnmente designadas por T, PT, CT, y CPT. Esto (en especial la violación CPT) parece estar más allá del poder de los métodos convencionales de cuantización.

cuantizadores de la gravitación necesitarán tratar de buscar en otra parte la "explicación" del bajo valor de la entropía en el *big bang*.

Quizá muchos físicos sostendrían que una hipótesis tal como la de una curvatura de Weyl inicialmente nula, al ser una elección de "condición de contorno" y no una ley dinámica, no es algo cuya explicación esté dentro de los poderes de la física. En realidad, están alegando que estamos ante un "acto divino", y no somos nosotros nadie para intentar comprender por qué nos ha sido dada una condición de contorno antes que otra. Sin embargo, como hemos visto, la limitación que esta hipótesis ha impuesto en la "aguja del Creador" no es menos extraordinaria ni menos precisa que toda la notable y delicada coreografía organizada que constituyen las leyes dinámicas que hemos llegado a comprender a través de las ecuaciones de Newton, Maxwell, Einstein, Schrödinger, Dirac y otros. Aunque pueda parecer que la segunda ley de la termodinámica tiene un carácter vago y estadístico, ella surge de una ligadura geométrica de la más suprema precisión. Me parece poco razonable desesperar de obtener una comprensión científica de las ligaduras que son operativas en la "condición de contorno" que era el *big bang* cuando la aproximación científica ha demostrado ser tan valiosa para la comprensión de las ecuaciones dinámicas. Para mi modo de pensar, lo primero forma parte de la ciencia tanto como lo segundo, aunque una parte de la ciencia que por ahora no entendemos adecuadamente.

La historia de la ciencia nos ha enseñado lo valiosa que ha sido la idea según la cual las *ecuaciones dinámicas* de la física (leyes de Newton, ecuaciones de Maxwell, etc.) han sido separadas de las llamadas *condiciones de contorno;* condiciones que hay que imponer para que las soluciones apropiadas de estas ecuaciones sean seleccionadas de entre el laberinto de las soluciones inapropiadas. Históricamente han sido las ecuaciones dinámicas las que han resultado tener formas simples. Los movimientos de las partículas satisfacen leyes sencillas pero, muy a menudo, las *configuraciones reales* de las partículas que resultan en el Universo no parecen serlo. A veces tales disposiciones parecen sencillas a primera vista —como es el caso de las órbitas elípticas del movimiento planetario como las concibió Kepler—pero luego se encuentra que su simplicidad es una *consecuencia* de las leyes dinámicas. La comprensión más profunda ha derivado siempre de las leyes dinámicas, y tales disposiciones simples tienden también a ser simples aproximaciones otras mucho más complicadas, tales como los movimientos planetarios perturbados (no muy elípticos) que se observan realmente, siendo éstos explicados a partir de las ecuaciones dinámicas de Newton. Las condiciones de contorno sirven para "arrancar" el sistema en cuestión y las ecuaciones dinámicas lo llevan en adelante. Una de las realizaciones más importantes de la ciencia física es que podemos separar el comportamiento dinámico de la cuestión de la disposición de los contenidos reales del Universo.

He dicho que esta separación en ecuaciones dinámicas y condiciones de contorno ha sido históricamente de vital importancia. El hecho de que sea posible hacer esta separación es una propiedad del tipo *particular* de ecuaciones (ecuaciones diferenciales) que parecen darse siempre en la física. Pero no creo que ésta sea una división que vaya a permanecer. En mi opinión, cuando finalmente lleguemos a comprender las leyes, o principios, que gobiernan *realmente* el comportamiento de nuestro Universo —en lugar de esas maravillosas aproximaciones que hemos llegado a comprender, y que constituyen nuestras teorías SUPREMAS hasta la fecha— encontraremos que esta distinción entre ecuaciones dinámicas y condiciones de contorno desaparecerá. En su lugar, habrá sólo un maravilloso esquema consistente y general. Por supuesto que al decir esto estoy expresando una opinión muy personal; muchos otros podrían no estar de acuerdo con ella. Pero es un punto de vista como éste el que tengo vagamente en la

cabeza al tratar de explorar las implicaciones de alguna teoría desconocida de la gravitación cuántica. (Este punto de vista afectará también a algunas de las consideraciones más especulativas del capítulo final.)

¿Cómo explorar las implicaciones de una teoría desconocida? Quizá las cosas no sean en absoluto tan desesperadas como pueda parecer. ¡Consistencia es la clave! En primer lugar, estoy pidiendo al lector que acepte que nuestra teoría supuesta —que llamaré GQC ("gravitación cuántica correcta")— proporcionará una explicación de la hipótesis de curvatura de Weyl (HCW). Esto significa que las singularidades *iniciales* deben estar limitadas de modo que WEYL = O en el futuro inmediato de la singularidad. Esta ligadura existirá como consecuencia de las leyes de la GQC, y por lo tanto debe aplicarse a *cualquier* "singularidad inicial", no sólo a la singularidad concreta que conocemos como *big bang.* No estoy diciendo que sea necesario que *haya* singularidades iniciales en nuestro Universo real además del *big bang,* sino que la idea es que, si las hubiera cualquiera de ellas tendría que estar limitada por la HCW. Una singularidad inicial sería una *de* la cual, en principio, podrían proceder las partículas. Este es el comportamiento contrario al de las singularidades de los agujeros negros, que son singularidades *finales,* en las que pueden caer las partículas.

Un tipo posible de singularidad inicial distinta de la del *big bang* sería la singularidad en un *agujero blanco* que, como recordamos en el capítulo VII, es el inverso temporal de un agujero negro (recuérdese la fig. VII. 14). Pero hemos visto que las singularidades en el interior de los agujeros negros satisfacen WEYL $\to \infty$, de modo que para un agujero blanco también debemos tener WEYL $\to \infty$. Pero la singularidad es ahora una singularidad *inicial,* para la que la HCW requiere WEYL = 0. En consecuencia, la HCW *descarta* la ocurrencia de agujeros blancos en nuestro Universo. (Afortunadamente esto no sólo es deseable sobre bases termodinámicas —pues los agujeros blancos desobedecerían descaradamente la segunda ley de la termodinámica—, sino que también es consistente con las observaciones. De cuando en cuando, diversos astrofísicos han postulado la existencia *de* agujeros blancos para intentar explicar ciertos fenómenos, pero esto plantea siempre más problemas de los que resuelve.) Nótese que no estoy calificando al propio *big bang* como "agujero blanco". Una agujero blanco poseería una singularidad inicial *localizada* que no podría satisfacer WEYL = 0; pero el *big bang* omnicomprensivo *puede* tener WEYL = 0 y la HCW le permite existir con tal de que tenga esta ligadura.

Existe otro tipo de posibilidad para una "singularidad inicial", a saber: el mismo punto de *explosión de un agujero negro* que finalmente ha *desaparecido* tras los (pongamos por caso) 10^{64} años de evaporación de Hawking . Hay muchas especulaciones sobre la naturaleza exacta de este supuesto (argumentado muy plausiblemente) fenómeno. Pienso que es probable que no haya aquí conflicto con la HCW. Semejante explosión (localizada) podría ser en efecto instantánea y simétrica, y no veo conflicto con la hipótesis WEYL = 0. (En cualquier caso, suponiendo que no existan miniagujeros negros es probable que la primera de tales explosiones no tenga lugar antes de que el Universo alcance un tiempo de existencia 10^{54} veces mayor que el tiempo T que lleva existiendo ahora. Para hacernos una idea de lo que significa 10^{54} x T, pensemos que si T se comprimiera hasta el más corto intervalo que puede ser medido —el más pequeño de los tiempos de desintegración de cualquier partícula inestable— entonces nuestra edad *real* del Universo presente, en esta escala, se quedaría corta respecto a 10^{54} x T ¡en un factor de un billón! Algunos

siguen una línea diferente de la que estoy proponiendo. Ellos argumentarán [3] que la GQC no debería tener asimetría temporal sino que, de hecho, permitiría *dos* tipos de estructura para las singularidades, uno de los cuales exige WEYL = 0, y el otro que permite WEYL $\to \infty$ .Habría una singularidad del primer tipo en nuestro Universo, y nuestra percepción de la dirección del tiempo (debido a la segunda ley subsiguiente) es tal que sitúa esta singularidad en lo que llamamos el "pasado" más que en lo que llamamos el "futuro". Sin embargo, me parece que este razonamiento no es adecuado tal como está. No explica por qué no hay *otras* singularidades iniciales del tipo WEYL $\to \infty$ (ni otras del tipo WEYL = 0). ¿Por qué, según esta idea, el Universo no está plagado de agujeros blancos? Puesto que presumiblemente *está* plagado de agujeros *negros* necesitamos una explicación de por qué no existen los blancos.*

Otro argumento que se invoca a veces en este contexto es el llamado *principio antrópico (cfr.* Barrow y Tipler, 1986). Según este razonamiento, el Universo particular en que nosotros mismos nos observamos habitar está seleccionado de entre todos los universos *posibles* por el hecho de que es necesario que *nosotros* (o, al menos, alguna especie de criatura viviente) estemos presentes para observarlo. (Discutiré otra vez el principio antrópico en el capítulo X.) Se afirma, utilizando este argumento, que los seres inteligentes sólo podrían habitar un universo con un tipo de *big bang* muy especial y, por lo tanto, algo semejante a la HCW podría ser una consecuencia de este principio. Sin embargo, el argumento nunca puede aproximarse a la cifra requerida de $10^{10^{123}}$, para la "particularidad" del *big bang* a que llegábamos en el capítulo VII. Mediante un cálculo aproximado se puede ver que todo el Sistema Solar junto con todos sus habitantes pudo ser creado sencillamente a partir de colisiones aleatorias de partículas de un modo más "barato" que éste, a saber: con una "improbabilidad" (medida en términos de volumen del espacio de fases) de "sólo" una parte en mucho menos de $10^{10^{60}}$. Esto es todo lo que el principio antrópico puede hacer por nosotros, y aún nos quedamos enormemente cortos respecto a la cifra requerida. Además, como sucedía con el punto de vista discutido inmediatamente antes este principio antrópico no ofrece explicación para la ausencia de agujeros blancos.

ASIMETRÍA TEMPORAL EN LA REDUCCIÓN DEL VECTOR DE ESTADO

Parece que tenemos que aceptar la conclusión de que la GQC debe ser una teoría con asimetría temporal, en donde la HCW (o algo muy parecido) sea una de las consecuencias de la teoría. ¿Como es posible obtener una teoría con asimetría temporal a partir de dos ingredientes con simetría temporal: la teoría cuántica y la relatividad general? Existen según parece, algunas posibilidades técnicas imaginables para conseguirlo, ninguna de las cuales ha sido explorada en profundidad *(cfr.* Ashtekar *et al,* 1989). Sin embargo, deseo explorar una línea diferente. He indicado que la teoría cuántica es "tiempo-simétrica", pero en realidad esto sólo se aplica a la

---

[3] Por lo que puedo deducir, un punto de vista de este tipo está implícito en las propuestas actuales de Hawking para una explicación de estas cuestiones por medio de la gravitación cuántica (Hawking, 1987, 1988). Una propuesta de Hartle y Hawking (1983) de un origen cuántico-gravitatorio para el estado inicial es posiblemente el tipo de cosa que *podría* dar sustancia teórica a una condición inicial del tipo WEYL = O, pero en estas ideas no está ausente ni mucho menos (en mi opinión) un input *esencial* con asimetría temporal.

* Alguien podría argumentar (correctamente) que las observaciones no son suficientemente claras ni mucho menos, para apoyar mi afirmación de que existen agujeros negros Pero no agujeros blancos en el Universo. Sin embargo, mi argumento es básicamente teórico. Los agujeros negros están de acuerdo con la segunda ley de la termodinámica, pero los agujeros blancos no lo están. (Por supuesto, se podría *postular* sencillamente la segunda ley y la ausencia de agujeros blancos, pero aquí estamos intentando penetrar más profundamente en los orígenes de la segunda ley.)

parte U de la teoría (ecuación de Schrödinger, etc.). En mis discusiones de la simetría temporal de las leyes físicas, al principio del capítulo VII, he dejado fuera deliberadamente la parte R (colapso de la función de onda). Parece predominar la opinión de que también R debería tener simetría temporal. Quizá esta opinión sea debida en parte a una hostilidad a considerar R como un "proceso" real independiente de U, de modo que la simetría temporal de U debería implicar la simetría temporal también para R. Deseo argumentar que esto *no es así:* R es tiempo-asimétrica, al menos si consideramos simplemente que "R" significa el procedimiento que adoptan realmente los físicos cuando calculan probabilidades en mecánica cuántica.

Permítaseme primero recordar al lector el procedimiento que se aplica en mecánica cuántica denominado reducción del vector de estado (R) (recuérdese la fig. VI.23). En la fig. VIII. 1 he indicado esquemáticamente la extraña manera en que se considera que evoluciona el vector de estado $|\psi\rangle$ en mecánica cuántica. En su mayor parte esta evolución se considera que procede según la evolución *unitaria* U (ecuación de Schrödinger), pero en instantes diversos, cuando se estima que ha tenido lugar una "observación" (o "medida"), se adopta el procedimiento R y el vector de estado $|\psi\rangle$ *salta* a otro vector de estado, por ejemplo $|\chi\rangle$, donde $|\chi\rangle$ es una entre dos o más posibilidades $|\chi\rangle$, $|\phi\rangle$, $|\theta\rangle$, ...que vienen determinadas por la naturaleza de la observación concreta O, que se está llevando a cabo. Ahora bien, la probabilidad p de saltar de $|\psi\rangle$ a $|\chi\rangle$ viene dada por la cantidad en que se reduce la longitud al cuadrado $|\psi|^2$ de $|\psi\rangle$ cuando se proyecta $|\psi\rangle$ sobre la dirección de $|\chi\rangle$ (en el espacio de Hilbert). (Matemáticamente esto es lo mismo que la cantidad en que se reduciría $|\chi|^2$, si se proyectara $|\chi\rangle$ en la dirección de $|\psi\rangle$.)

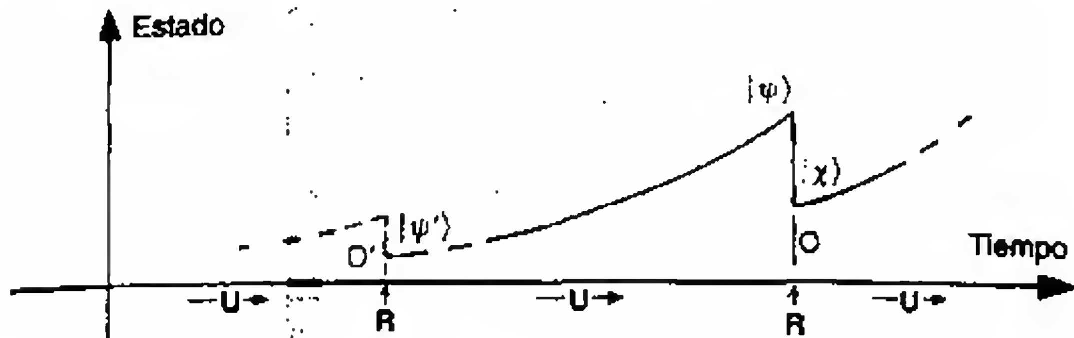


FIGURA VIII. 1. *Evolución temporal de un vector de estado: evolución unitaria suave U (según la ecuación de Schrödinger) interrumpida por la reducción discontinua R del vector de estado.*

Tal como está, este procedimiento tiene asimetría temporal puesto que inmediatamente *después* de que se haya hecho la observación O, el vector de estado es uno del *conjunto dado* $|\chi\rangle$, $|\phi\rangle$, $|\theta\rangle$,... *determinado por* O, mientras que inmediatamente *antes* de O, el vector de estado era $|\psi\rangle$, que *no* tiene por qué ser una de las opciones dadas. Sin embargo, esta asimetría es sólo aparente y puede remediarse adoptando un punto de vista diferente sobre la evolución del vector de estado. Consideremos una evolución mecánico-cuántica *invertida en el tiempo.* Esta excéntrica descripción se ilustra en la fig. VIII.2. Supongamos ahora que el estado es $|\chi\rangle$ inmediatamente

antes de O, en lugar de inmediatamente después, y supongamos que la evolución unitaria se aplica *hacia atrás en el tiempo* hasta el instante de la observación *previa* O'. Supongamos que este estado evolucionado hacia atrás se transforma en $|\chi'\rangle$ (inmediatamente en el futuro de la observación O')- En la descripción normal evolucionada hacia adelante de la fig. VIII.1 teníamos algún otro estado $|\psi'\rangle$ en el futuro inmediato de O' (el resultado de la observación de O', donde $|\psi'\rangle$ evolucionaría hacia delante hasta $|\psi\rangle$ en O en la descripción normal). Ahora, en nuestra descripción *invertida,* el vector de estado $|\psi'\rangle$ también tiene un papel: representa el estado del sistema en el *pasado* inmediato de O'. El vector de estado $|\psi'\rangle$ es el estado que se observaba realmente en O', de modo que en nuestro punto de vista con evolución hacia atrás consideramos que el estado $|\psi\rangle$ es el "resultado", en el sentido *inverso en el tiempo,* de la observación en O'. El cálculo de la probabilidad cuántica p' que relaciona el resultado de la observación en O' con la observación en O, viene dado ahora por la cantidad en que decrece $|\chi|^2$ al proyectar $|\chi'\rangle$ sobre la dirección de $|\psi'\rangle$ (siendo esta la misma cantidad en la que decrece $|\psi'|^2$ cuando se proyecta $|\psi'\rangle$ sobre la dirección de $|\chi'\rangle$)- Es una propiedad fundamental de la operación de U, el que efectivamente éste es exactamente el mismo valor que teníamos antes.[4]

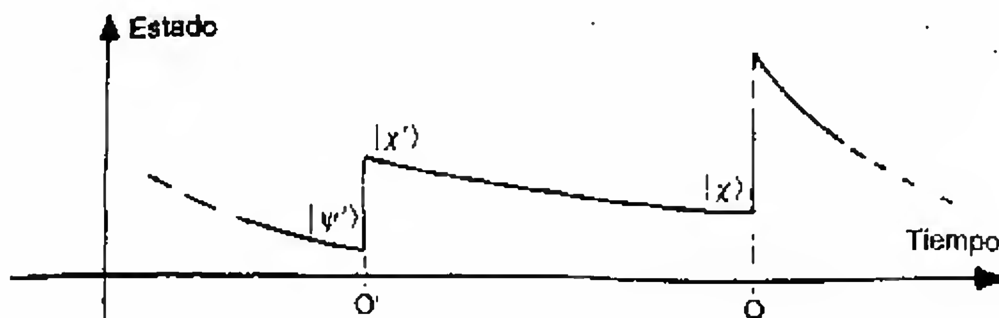


FIGURA VIII.2. *Una imagen más excéntrica de la evolución del vector de estado, en la que se utiliza una descripción invertida en el tiempo. La probabilidad calculada que relaciona la observación en O con la observación en O' sería la misma que en la fig- VIII. 1, pero ¿a que se refiere ahora este valor calculado?*

Parecería, por lo tanto, que hemos establecido que la *teoría cuántica posee simetría temporal,* incluso cuando tenemos en cuenta el proceso discontinuo descrito por la reducción del vector de estado R, además de la evolución unitaria ordinaria U. Sin embargo, *no es así.* Lo que describe la probabilidad cuántica *p* — de cualquier forma que se calcule — es la probabilidad de encontrar

[4] Estos hechos son algo más transparentes en términos de la operación del *producto escalar* $\langle\psi|\chi\rangle$ dada en la nota 6 del capítulo VI. En la descripción hacia delante en el tiempo calculamos la probabilidad *p* mediante

$$p = |\langle\psi|\chi\rangle|^2 = |\langle\chi|\psi\rangle|^2$$

y en la descripción hacia atrás, mediante

$$p = |\langle\chi'|\psi'\rangle|^2 = |\langle\psi'|\chi'\rangle|^2$$

El hecho de que éstas sean iguales se sigue de $\langle\psi'|\chi'\rangle = \langle\chi'|\psi'\rangle$ que es lo que esencialmente entendemos por "evolución unitaria".

el resultado (a saber $|\chi\rangle$) en O *dado* el resultado (a saber $|\psi'\rangle$) en O'. Esta no es necesariamente la misma que la probabilidad del resultado en O' *dado* el resultado en O. La última [5] sería realmente la que estaría obteniendo nuestra mecánica cuántica invertida en el tiempo. Es curioso cuántos físicos parecen suponer tácitamente que estas dos probabilidades son iguales. (Yo mismo he sido culpable de esta presunción, *cfr.* Penrose 1979b, p. 584.) Sin embargo, es probable que estas dos probabilidades sean completamente diferentes de hecho y solo la primera está dada correctamente por la mecánica cuántica.

Veamos esto en un caso concreto muy simple. Supongamos que tenemos una lámpara L y una fotocélula (es decir, un detector de fotones) F. Entre L y F tenemos un espejo semirreflectante E que está inclinado un cierto ángulo, pongamos que 45°, respecto a la recta que une L y F (véase fig. VIII.3). Supongamos que la lámpara emite un fotón de cuando en cuando, de un modo aleatorio, y que la lámpara está construida (se podrían utilizar espejos parabólicos) de modo que estos fotones se dirigen exactamente hacia F. Siempre que la fotocélula reciba un fotón registrará este hecho, y suponemos que tiene una fiabilidad del 100%. Podemos suponer también que siempre que se emita un fotón, este hecho se *registra* en L, de nuevo con una confiabilidad del 100%. (No hay conflicto con los principios de la mecánica cuántica en ninguno de estos requisitos ideales, aunque podría haber dificultades para conseguir en la práctica esta eficiencia.) El espejo semirreflectante E es tal que refleja exactamente la mitad de los fotones que le llegan y transmite la otra mitad. La función de onda del fotón incide en el espejo y se desdobla en dos. Existe una amplitud $1/\sqrt{2}$ para la parte reflejada de la onda y $1/\sqrt{2}$ para la parte transmitida. Ambas partes deben considerarse "coexistentes" (en la descripción normal hacia adelante en el tiempo) hasta el instante en que se estima que se ha realizado una "observación". En este punto, estas alternativas coexistentes se resuelven por sí mismas en opciones *reales* —una *o* la otra— con probabilidades dadas por los cuadrados de (los módulos de) estas amplitudes, a saber $(1/\sqrt{2})^2 = 1/2$, en cada caso. Cuando se ha hecho la observación, resulta que las probabilidades de que el fotón sea reflejado o transmitido han sido ambas realmente *de* 1/2.

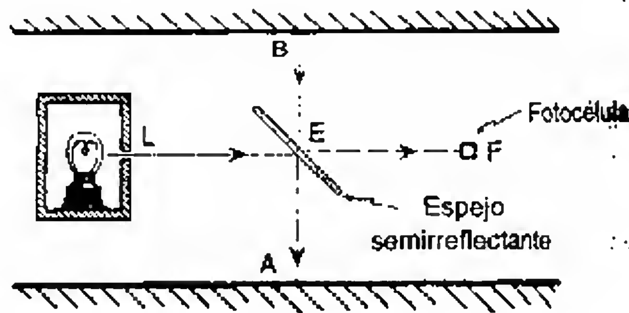


FIGURA VIII.3. *Irreversibilidad temporal de R en un experimento cuántico sencillo. La probabilidad de que la fotocélula detecte un fotón,* dado *que la fuente ha emitido uno, es exactamente 1/2; pero la probabilidad de que la fuente haya emitido un fotón,* dado *que la fotocélula detecta uno, no es ciertamente 1/2.*

[5] Algunos lectores pueden tener dificultad en comprender lo que puedo querer decir al preguntar cuál es la probabilidad de un suceso pasado dado un suceso futuro. Sin embargo no hay un problema esencial en esto. Imaginemos toda la historia del Universo representada en el espacio-tiempo. Para encontrar la probabilidad de que ocurra p, dado que ha ocurrido *q,* imaginémonos examinando todas las ocurrencias de *q y* contando qué fracciones de estas están acompañadas de *p.* Esta es la probabilidad buscada. No importa si *q* es el tipo de suceso que normalmente ocurriría más tarde o antes que *p.*

Veamos como se aplica esto a nuestro experimento real. Supongamos que se registra que L ha emitido un fotón. La función de onda del fotón se desdobla en el espejo y llega a F con amplitud $1/\sqrt{2}$, de modo que la fotocélula registra o no, con una probabilidad 1/2 para cada caso. La otra parte de la función de onda llega a un punto A en la *pared del laboratorio* (véase fig. VIII.3), de nuevo con amplitud $1/\sqrt{2}$. Si F *no* registra, entonces debe considerarse que el fotón ha incidido en la pared en A, pues si hubiéramos colocado otra fotocélula en A entonces esta nueva fotocélula registraría siempre que no lo hiciera la fotocélula en F —suponiendo que L ha registrado realmente la emisión de un fotón— y no registraría siempre que registrara F. En este sentido no es necesario colocar una fotocélula en A. Podemos inferir lo que la fotocélula en A *habría* hecho, si hubiera estado allí, simplemente mirando en L y F.

Estaría claro cómo procede el cálculo mecánico-cuántico. Planteamos la pregunta:

"Dado que L registra, ¿cuál es la probabilidad de que registre F?"

Para responderla notemos que existe una amplitud $1/\sqrt{2}$ para que el fotón recorra el camino LEF y una amplitud $1/\sqrt{2}$ para que recorra el camino LEA. Elevando al cuadrado encontramos las probabilidades respectivas 1/2 y 1/2 de llegar a F y de llegar a A. La respuesta mecánico-cuántica a nuestra pregunta es por lo tanto

"un medio".

Esta es en realidad la respuesta que obtendríamos experimentalmente. Podríamos utilizar exactamente igual el excéntrico procedimiento con el "tiempo invertido" para obtener la misma respuesta. Supongamos que notamos que F ha registrado. Consideremos una función de onda con *el* "tiempo invertido" para el fotón, suponiendo que el fotón llega finalmente a F. A medida que seguimos la pista hacia atrás en el tiempo, el fotón retrocede desde F hasta que *alcanza* el espejo E. En este punto la función de onda se bifurca y existe una amplitud $1/\sqrt{2}$ de que llegue a la lámpara L, y una amplitud $1/\sqrt{2}$ de que sea reflejada en E hasta llegar a *otro punto* en la pared del laboratorio, a saber, B en la fig. VIII.3. Elevando al cuadrado obtenemos de nuevo el valor 1/2 para las dos probabilidades. No obstante, debemos tener cuidado al señalar a qué preguntas responde estas respuestas. Existen las dos preguntas, "Dado que L registra cuál es la probabilidad de que F registre?", igual que antes, y la pregunta más excéntrica, "dado que el fotón es lanzado desde la pared en B", ¿cuál es la probabilidad de que F registre?"

Podemos considerar que, en cierto sentido, ambas respuestas son experimentalmente "correctas", aunque la segunda (lanzamiento desde la pared) sería una inferencia más que el resultado de una serie *real* de experimentos. Sin embargo, ninguna de estas preguntas es *la inversa en el tiempo* de la pregunta que planteábamos antes. Esta sería:

"Dado que F registra, ¿cuál es la probabilidad de que L registre?"

Notemos que la respuesta experimental *correcta* a esta pregunta no es en absoluto "un medio" sino

"uno".

Si la fotocélula registra, realmente entonces es virtualmente cierto que el fotón procede de la *lámpara* y no de la pared del laboratorio. En el caso de nuestra pregunta invertida en el tiempo, el cálculo mecánico-cuántico nos ha dado *una respuesta totalmente errónea.*

La consecuencia de esto es que las reglas para la parte R de la mecánica cuántica no pueden utilizarse simplemente para estas preguntas con inversión temporal. Si queremos calcular la probabilidad de un estado *pasado* sobre la base de un estado *futuro* conocido, obtendremos respuestas completamente erróneas si tratamos de adoptar el procedimiento estándar R consistente en tomar la amplitud mecánico-cuántica y elevar su módulo al cuadrado. Este procedimiento sólo funciona para calcular las probabilidades de estados *futuros* sobre la base de estados *pasados,* y funciona de manera soberbia. Me parece evidente que, sobre esta base, el procedimiento R *no puede tener simetría temporal* (y, dicho sea de paso, no puede haber por consiguiente una deducción a partir del procedimiento con simetría temporal U).

Mucha gente podría adoptar la postura de que la razón de esta discrepancia con la simetría temporal es que de algún modo la segunda ley de la termodinámica se ha colado en el argumento, introduciendo una asimetría temporal añadida que no está descrita por el procedimiento de tomar el cuadrado de la amplitud. De hecho, parece ser cierto que cualquier dispositivo físico de medida capaz de llevar a cabo el procedimiento R debe involucrar una "irreversibilidad termodinámica", de modo que la entropía aumenta siempre que tiene lugar una medida. Creo que es probable que la segunda ley *esté* involucrada de un modo esencial en el proceso de medida. Además, no parece de mucho sentido físico tratar de invertir el tiempo en la operación *completa* de un experimento mecánico-cuántico como el experimento (idealizado) descrito arriba, incluyendo el registro de todas las medidas implicadas. No he tratado la cuestión de hasta dónde podemos ir realmente con la inversión temporal de un experimento; he considerado sólo la aplicabilidad de ese curioso procedimiento mecánico-cuántico que obtiene probabilidades correctas elevando al cuadrado los módulos de las amplitudes. Es un hecho sorprendente que este simple procedimiento pueda aplicarse en la dirección del futuro sin que sea necesario invocar ningún otro conocimiento de un sistema. De hecho, forma parte de la teoría el que *no podamos* influir en estas probabilidades: las probabilidades de la teoría cuántica son completamente *estocásticas.* Sin embargo, si intentamos aplicar estos procedimientos en la dirección del pasado (es decir, para retrodicción más que para predicción) entonces nos equivocamos completamente. Deben invocarse cualquier número de excusas, circunstancias atenuantes u otros factores para explicar *por qué* el procedimiento de tomar el cuadrado de las amplitudes no se aplica correctamente en la dirección del pasado, pero sigue en pie el hecho de que no lo hace. Estas excusas son sencillamente innecesarias en la dirección del futuro. El procedimiento R, *tal como se utiliza realmente, no* es simétrico respecto al tiempo.

LA CAJA DE HAWKING: ¿UNA CONEXIÓN CON LA HIPÓTESIS DE CURVATURA DE WEYL?

Está bien, estará pensando sin duda el lector, pero ¿qué tiene todo esto que ver con la HCW o la GQC? Es cierto que la *segunda ley,* como hoy opera, puede perfectamente ser parte de la operación de R, pero ¿dónde hay algún papel destacable para las singularidades del espacio-tiempo o la gravitación cuántica en estas continuas ocurrencias "cotidianas" de reducción del vector de estado? Para abordar esta cuestión deseo describir un espectacular "experimento mental" propuesto originalmente por Stephen Hawking, si bien el propósito con el que se expone no forma parte de lo que Hawking pretendió originalmente.

Imaginemos una caja sellada de proporciones monstruosas. Se considera que sus paredes son totalmente reflectantes e impermeables a cualquier influencia. Ningún objeto material puede

atravesarlas, ni tampoco ninguna señal electromagnética, o neutrino, o ninguna otra cosa. Todo debe ser reflejado, ya incida desde dentro o desde fuera. Incluso está prohibido que la atraviesen los efectos de la gravitación. No existe ninguna sustancia real de la que puedan construirse estas paredes. Nadie podría *realizar* realmente el "experimento" que voy a describir. (Ni nadie querría hacerlo, como vamos a ver) Este no es el punto. En un experimento mental se trata de descubrir principios generales a partir de la simple consideración mental de experimentos que *se podrían* realizar. Se ignoran las dificultades tecnológicas siempre que estas dificultades no se deriven de los principios generales en consideración. (Recordemos la discusión del gato de Schrödinger en el capítulo VI.) En nuestro caso, las dificultades para construir las paredes de nuestra caja deben considerarse, para este propósito, como puramente "tecnológicas", de modo que estas dificultades serán ignoradas. En el interior de la caja hay una gran cantidad de sustancia material de algún tipo. No importa mucho cuál sea esta sustancia. Sólo nos interesa su masa total M, que deberá ser muy grande, y el gran volumen V de la caja que la contiene. ¿Qué vamos a hacer con nuestra caja construida con tanto gasto y con su contenido de tan nulo interés? El experimento va a ser lo más aburrido que se pueda imaginar. Vamos a dejarla aislada para siempre.

La cuestión que nos interesa es el destino final del contenido de la caja. Según la segunda ley de la termodinámica su entropía deberá crecer. La entropía aumentará hasta que se alcance su valor máximo, de modo que el material habrá llegado al "equilibrio térmico". Nada sucedería a partir de entonces, si no fuera por las "fluctuaciones" en las que se alcanzan temporalmente desviaciones (relativamente) breves respecto al equilibrio térmico. En nuestra situación, suponemos que M es suficientemente grande, y V es algo apropiado (*muy* grande, pero no demasiado grande), de modo que cuando se alcance el "equilibrio térmico" la mayor parte del material habrá colapsado en un *agujero negro,* con sólo un poco de materia y radiación rodeándolo, que constituye un llamado "baño térmico" (¡muy frío!) en el que está inmerso el agujero negro. Para ser precisos, podríamos escoger M como la masa del Sistema Solar y V como el tamaño de la Vía Láctea. Entonces la temperatura del "baño" sería sólo de unos 10^{-7} grados por encima del cero absoluto.

Para comprender más claramente la naturaleza de este equilibrio y estas fluctuaciones recordemos el concepto de *espacio de fases* que ya encontramos en los capítulos V y VII, especialmente en relación con la definición de entropía. La fig. VIII.4 da una descripción esquemática de todo el espacio de fases P del contenido de la caja de Hawking. Como recordamos, un espacio de fases es un espacio de muchas dimensiones, cada uno de cuyos puntos representa todo un posible estado global del sistema en consideración; en nuestro caso, el contenido de la caja. Así, cada punto de P codifica las posiciones y momentos de todas las partículas presentes en la caja, además de toda la información necesaria sobre la *geometría espacio-temporal* dentro de la caja. La subregión B (de P) a la derecha de la fig. VIII.4 representa la totalidad de los estados en los que hay un *agujero negro* dentro de la caja (incluyendo todos los casos en los que hay más de un agujero negro), mientras que la subregión A de la izquierda representa la totalidad de los estados libres de agujeros negros.

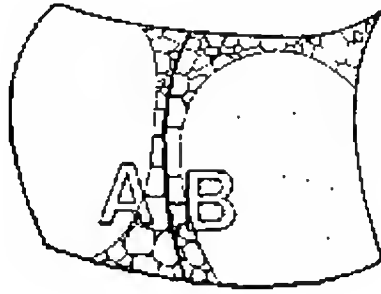


FIGURA VIII.4. *El espacio de fases P de la caja de Hawking. La región A corresponde a las situaciones en las que* no *existe agujero negro en la caja, y Balas situaciones en las que* sí *existe un agujero negro (o mas de uno) en la caja.*

Debemos suponer que cada una de las regiones A y B está además subdividida en compartimentos más pequeños de acuerdo con la división de "grano grueso" que es necesaria para la definición precisa de la entropía (*cfr.* fig. VII.3), pero estos detalles no nos interesan aquí. Todo lo que tenemos que señalar en esta etapa es que el mayor de estos compartimentos —que representa el equilibrio térmico, con presencia de un agujero negro— es la porción mayor de B, mientras la mayor porción de A (algo más pequeña) es el compartimento que representa lo que *aparenta* ser equilibrio térmico, salvo el hecho de que no hay agujero negro presente.

Recordemos que en cualquier espacio de fases hay un campo de flechas (campo vectorial) que representa la evolución temporal del sistema físico (véase capítulo V ; también fig. V.I 1). Así, para ver qué sucederá a continuación con nuestro sistema seguimos simplemente a lo largo de las flechas en P (véase fig. VIII.5). Algunas de estas flechas pasan de la región A a la región B. Esto ocurre cuando se forma por primera vez un agujero negro por el colapso gravitatorio de materia. ¿Existen flechas que pasen de la región B a la región A? Sí que existen, pero sólo si tenemos en cuenta el fenómeno de la *evaporación de Hawking* a que aludimos antes . Según la teoría estrictamente *clásica* de la relatividad general, los agujeros negros sólo pueden engullir cosas; no pueden emitir cosas. Sin embargo, teniendo en cuenta los efectos mecánico-cuánticos, Hawking (1975) pudo demostrar que, después de todo, los agujeros negros deberían emitir cosas en el nivel cuántico, de

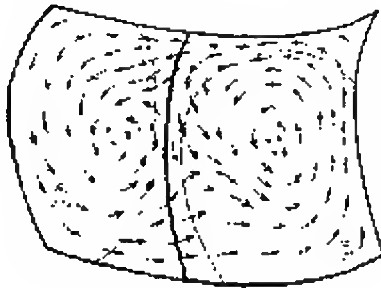


FIGURA VIII.5. *El "flujo hamiltoniano" del contenido de la caja de Hawking (compárese con fig. V. 11). Las líneas de flujo que cruzan de A a B representan el colapso en un agujero negro, y las que cruzan de B a A representan la desaparición de un agujero negro por evaporación.*

acuerdo con el proceso de *radiación de Hawking.* (Esto ocurre por vía del proceso cuántico de "creación de pares virtuales", por el que continuamente se están creando partículas y antipartículas a partir del vacío —por un breve instante— normalmente para aniquilarse entre sí inmediatamente después sin dejar huella. Cuando está presente un agujero negro, sin embargo, puede "engullir" una de las partículas del par antes de que tenga tiempo de ocurrir la aniquilación, y su compañera puede escapar del agujero. Estas partículas que escapan constituyen la radiación de Hawking.) En el curso normal de las cosas, esta radiación de Hawking es realmente pequeñísima. Pero en el estado de equilibrio térmico la cantidad de energía que pierde el agujero negro por radiación de Hawking compensa exactamente la energía que gana tragando otras "partículas térmicas" que se están moviendo por el "baño térmico" en el que se encuentra el propio agujero negro. En ocasiones, debido a una "fluctuación", el agujero podría emitir algo más o engullir un poco menos y, en consecuencia, perder energía. Al perder energía pierde masa (por la ecuación de Einstein $E = mc^2$) *y,* según las reglas que gobiernan la radiación de Hawking, se hace un poco más caliente. *Muy* pero muy ocasionalmente, cuando la fluctuación es suficientemente grande, es incluso posible que el agujero negro entre en una situación incontrolada en la que se va haciendo cada vez más caliente, perdiendo cada vez más energía a medida que esto sucede, y haciéndose cada vez más pequeño hasta que finalmente (es de suponer) desaparece completamente en una violenta explosión. Cuando esto sucede (y suponiendo que no haya otros agujeros en la caja) tenemos la situación en la que, en nuestro espacio de fases P, pasamos de la región B a la región A, de modo que realmente hay flechas de B a A.

En este punto haré un comentario sobre lo que significa una "fluctuación". Recordemos los compartimentos de grano-grueso que consideramos en el capítulo anterior. Los puntos del espacio de fases que pertenecen a un mismo compartimento deben considerarse "indistinguibles" (macroscópicamente) unos de otros. La entropía se incrementa debido a que al seguir las flechas tendemos a entrar en compartimentos cada vez más enormes a medida que *avanza* el tiempo. Finalmente, el punto en el espacio de fases se pierde en el compartimento más enorme de todos, a saber, el que corresponde al equilibrio térmico (entropía máxima). Sin embargo, esto sólo será verdadero hasta cierto punto. Si esperamos lo suficiente, el punto en el espacio de fases podrá encontrar *eventualmente* un compartimento más pequeño y en consecuencia la entropía descenderá. Normalmente esto no durará mucho (hablando en términos relativos) y la entropía pronto volverá a crecer cuando el punto del espacio de fases entre de nuevo en el compartimento más grande. Esto es una *fluctuación,* con su momentánea disminución de entropía. Normalmente el descenso de la entropía no es muy grande, pero muy de tarde en tarde ocurrirá una *enorme* fluctuación y la entropía podrá descender de forma sustancial y quizá permanecer con un bajo valor durante un intervalo de tiempo considerable.

Este es el tipo de cosas que necesitamos para ir de la región B a la región A vía el proceso de evaporación de Hawking. Se necesita una fluctuación muy grande debido a que debe atravesarse un compartimento muy pequeño en el que las flechas pasan de B a A. Análogamente, cuando nuestro punto del espacio de fases yace en el interior del compartimento más grande de A (que representa el estado de equilibrio térmico sin agujeros negros), deberá pasar mucho tiempo antes de que tenga lugar un colapso gravitatorio y el punto se mueva hasta B. Se necesita de nuevo una gran fluctuación. (La radiación térmica no está dispuesta a sufrir un colapso gravita torio.)

¿Es *mayor* el número de flechas que llevan de A a B que el de flechas que llevan de B a A, o es el *mismo* en ambos casos? Esto será un punto importante para nosotros. Para plantear la cuestión

de otra forma: ¿es "más fácil" para la naturaleza producir un agujero negro mediante el colapso gravitatorio de partículas térmicas, o lo es deshacerse de un agujero negro por radiación de Hawking, o ambas cosas son igualmente "difíciles"? Estrictamente hablando no es el "número" de flechas lo que nos interesa sino la velocidad del flujo del volumen del espacio de fases. Imaginemos el espacio de fases lleno de algún tipo de fluido incompresible (de alta dimensión). Las flechas representan el flujo de este fluido.

Recordemos el *teorema de Liouville* que se describió en el capítulo V. El teorema de Liouville afirma que el flujo conserva el volumen de1 espacio de fases, equivalente a decir que nuestro fluido en el espacio de fases es realmente incompresible. El teorema de Liouville parece decirnos que el flujo de A a B debe ser *igual* al flujo de B a A ya que, al ser incompresible el "fluido" del espacio de fases, no puede acumularse en una u otra parte. Parecería así que debe tener exactamente la misma "dificultad" construir un agujero negro a partir de la radiación térmica que destruirlo.

Esta fue en efecto la propia conclusión de Hawking, aunque él llegó a esta idea basándose en consideraciones diferentes. El principal argumento de Hawking era que toda la física básica implicada en el problema tiene *simetría temporal* (relatividad general, termodinámica, los procedimientos unitarios estándar de la teoría cuántica), de modo que si hiciéramos marchar el reloj hacia atrás obtendríamos la misma respuesta que si marchara hacia delante. Esto equivale simplemente a invertir las direcciones de todas las flechas en P. Entonces se seguiría también de *este* razonamiento que debe haber exactamente tantas flechas de A a B como de B a A *con tal de que* se verifique que el inverso temporal de la región B sea de nuevo la región B (y, de modo equivalente, que el inverso temporal de A sea de nuevo A). Esta condición equivale a la famosa sugerencia de Hawking de que los agujeros negros y sus inversos temporales, a saber, los agujeros blancos, son en realidad físicamente idénticos. Su razonamiento era que en una física con simetría temporal el estado de equilibrio térmico también debería tener simetría temporal. No quiero entrar aquí en una discusión detallada de esta sorprendente posibilidad. La idea de Hawking era que la radiación mecánico-cuántica de Hawking podía considerarse de algún modo como la inversa temporal del "engullimiento" clásico de material por el agujero negro. Su hipótesis, aunque ingeniosa, implica diversas dificultades teóricas y personalmente no creo que pueda funcionar.

En cualquier caso, esta hipótesis no es compatible con las ideas que estoy desarrollando aquí. He argumentado que, mientras que los agujeros negros deberían existir, los agujeros blancos están "prohibidos" por la *hipótesis de curvatura de Weyl.* La HCW introduce en la discusión una *asimetría temporal* que no era considerada por Hawking. Debería señalarse que puesto que los agujeros negros y sus singularidades en el espacio-tiempo son una parte importante de la discusión de lo que sucede en el interior de la caja de Hawking, la física desconocida que debe gobernar el comportamiento de tales singularidades está ciertamente involucrada. Hawking adopta la postura de que esta física desconocida debería ser una teoría de la gravitación cuántica *con simetría temporal,* mientras que yo sostengo que es la GQC con *asimetría* temporal. Estoy defendiendo que una de las implicaciones principales de la GQC debería ser la HCW (y, por consiguiente, la segunda ley de la termodinámica en la forma que la conocemos), de modo que trataremos de descubrir las implicaciones de la HCW para nuestro problema actual.

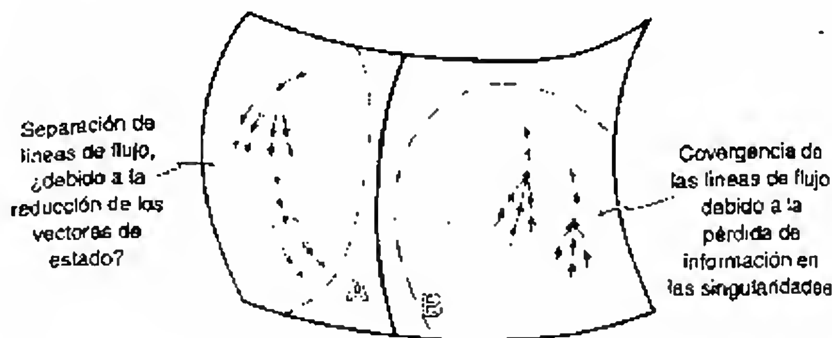


FIGURA VIII.6. *En la región B las líneas de flujo deben converger debido a la pérdida de información en las singularidades de los agujeros negros. ¿Se compensa esto con una creación de líneas de flujo debido al procedimiento cuántico R (principalmente en la región A)?*

Veamos cómo afecta la inclusión de la HCW a la determinación del flujo de nuestro "fluido incompresible" en P. En el espacio-tiempo, el efecto de la singularidad de un agujero negro es el de absorber y destruir toda la materia que incida en él. Lo que es más importante para nuestros propósitos presentes, es que *destruye información.* El efecto de esto, en P, es que algunas líneas de flujo convergerán (véase fig. VIII.6). Dos estados que antes eran diferentes pueden llegar a ser el mismo tan pronto como se destruya la información que era la distinción entre ellos. Cuando convergen las líneas de flujo en P tenemos una "violación" efectiva del teorema de Liouville. Nuestro "fluido" ya no es incompresible, sino que "se está *aniquilando continuamente"* dentro de la región B.

Parece que ahora estamos en dificultades. Si nuestro "fluido" se está destruyendo continuamente en la región B, entonces debe haber *más* líneas de flujo de A a B que de B a A, de modo que, después de todo, es "más fácil" crear un agujero negro que destruirlo. Esto podría tener sentido si no fuera por el hecho de que ahora sale más "fluido" de la región A del que entra. No hay agujeros negros en la región A —y los agujeros blancos están prohibidos por la HCW—, de modo que ciertamente el teorema de Liouville debería seguir siendo perfectamente válido en 1ª región A. Sin embargo, parece ahora que necesitamos algún medio de "crear fluido" en la región A para enmascarar la pérdida en la región B.

Que mecanismo puede haber para incrementar el número de líneas de flujo? Lo que necesitamos es que a veces un mismo estado pueda tener más de un posible sucesor (es decir, líneas de flujo que se bifurcan) Este tipo de incertidumbre en la evolución futura de un sistema físico lleva el "aroma" de la parte R de la teoría cuántica. ¿Sería posible que R sea, en cierto sentido, "la otra cara de la moneda" de la HCW? Mientras qué la HCW sirve para que converjan líneas de flujo en B, el procedimiento mecánico-cuántico R provoca que las líneas de flujo se bifurquen. Estoy afirmando de hecho que es un proceso *objetivo* mecano-cuántico de reducción del vector de estado (R) el que provoca que las líneas de flujo se bifurquen, y de este modo compensa exactamente la fusión de líneas de flujo debida a la HCW (fig. VIII.6).

Para que tenga lugar dicha bifurcación necesitamos que R tenga asimetría temporal, como ya hemos visto: recuérdese nuestro experimento anterior con la lámpara, la fotocélula y el espejo semirreflectante. Cuando la lámpara emite un fotón hay dos opciones (igualmente probables)

para el resultado final: o bien el fotón llega a la fotocélula y la fotocélula registra, o bien el fotón llega a la pared en A y la fotocélula no registra. En el espacio de fases para este experimento tenemos una línea de flujo que representa la emisión del fotón y esta línea se bifurca en dos: una que describe la situación en la que la fotocélula se dispara, y la otra que refleja la situación en que no lo hace. Esta parece ser una auténtica bifurcación debido a que sólo hay un estado de partida permitido y hay dos resultados posibles. El otro estado de partida que debimos considerar era la posibilidad de que el fotón fuera lanzado desde la pared del laboratorio en B, en cuyo caso habría dos estados de partida y dos resultados. Pero este estado de partida alternativo ha sido excluido sobre la base de la inconsistencia con la segunda ley de la termodinámica, es decir, desde el punto de vista aquí expresado, excluido finalmente por la HCW cuando se rastrea la evolución hacia el pasado.

Reiteraré que el punto de vista que estoy exponiendo no es realmente "convencional" —aunque no tengo completamente claro lo que diría un físico "convencional" para resolver las cuestiones que aquí se plantean. (Sospecho que no muchos de ellos han dedicado mucha atención a estos problemas.) Ciertamente he oído varias opiniones diferentes. Por ejemplo, algunos físicos han sugerido que la radiación de Hawking nunca podría dar lugar a que un agujero negro desaparezca *completamente,* sino que debería quedar siempre alguna pequeña "pepita". (Por lo tanto, en esta perspectiva, no hay flechas que vayan de B a A.) Esto apenas afecta a mi argumento (y, en realidad, lo refuerza). Podrían evitarse mis conclusiones, no obstante, postulando que el volumen total del espacio de fases P es realmente *infinito,* pero esto está en contra de ciertas ideas bastante básicas sobre la entropía de los agujeros negros y sobre la naturaleza del espacio de fases de un sistema (cuántico) acotado; y otras formas que he oído de evitar técnicamente mis conclusiones no me parecen más satisfactorias. Una objeción mucho más seria es la de que las idealizaciones implicadas en la construcción real de una caja de Hawking son demasiado grandes y se transgreden ciertas cuestiones de principio al suponer que se puede construir. Yo mismo no estoy muy seguro de ello, aunque estoy inclinado a creer que las idealizaciones necesarias pueden ser realmente admitidas.

Finalmente, hay un punto importante que quiero tratar. Empecé la discusión suponiendo que teníamos un espacio de fases *clásico,* y el teorema de Liouville se aplica a la física clásica. Pero entonces necesitaba ser considerado el fenómeno cuántico de la radiación de Hawking. (Y también se necesitaba la teoría cuántica para la *dimensionalidad finita* tanto como el volumen finito de P.) Como vimos en el capítulo VI, la versión cuántica del espacio de fases es el *espacio de Hilbert,* de modo que presumiblemente deberíamos haber utilizado a lo largo de toda nuestra discusión el espacio de Hilbert antes que el espacio de fases. En el espacio de Hilbert *existe* un teorema análogo al de Liouville; surge de lo que se llama la naturaleza *"unitaria"* de la evolución temporal U. Quizá todo mi argumento podría expresarse completamente en términos del espacio de Hilbert, en lugar del espacio de fases clásico, pero es difícil ver cómo discutir de esta forma el fenómeno clásico implícito en la geometría del espacio-tiempo de los agujeros negros. Mi opinión es que para la teoría *correcta* no serían apropiados ni el espacio de Hilbert ni el espacio de fases clásico, sino que tendríamos que utilizar algún tipo de espacio matemático hasta ahora no descubierto que fuera intermedio entre los dos. Por consiguiente, mi argumento debería considerarse sólo en un nivel heurístico, y es simplemente *sugerente* más que concluyente. De todas formas, creo que proporciona un fuerte motivo para pensar que la HCW y R están profundamente relacionados y que, en consecuencia, R *debe ser en realidad un efecto de gravitación cuántica.*

Para reiterar mis conclusiones: estoy desarrollando la hipótesis de que la reducción mecánico-cuántica del vector de estado es en realidad la otra cara de la moneda de la HCW. Según esta idea, las dos implicaciones principales de nuestra deseada teoría de la "gravitación cuántica correcta" (GQC) serán la HCW y R. El efecto de la HCW es la *confluencia* de líneas de flujo en el espacio de fases, mientras que el efecto de R es una *ramificación* de las líneas de flujo exactamente compensadora. Ambos procesos están íntimamente asociados a la segunda ley de la termodinámica.

Nótese que la confluencia de líneas de flujo tiene lugar siempre dentro de la región B, mientras que la *ramificación* de líneas de flujo, puede tener lugar en A o en B. Recordemos que A representa la *ausencia* de agujeros negros, así que la reducción del vector de estado puede tener lugar cuando no hay agujeros negros. Evidentemente no es necesario tener un agujero negro en el laboratorio para que R sea efectivo (como sucede en nuestro experimento con el fotón recién considerado). Aquí solamente estamos interesados con un balance global entre posibles cosas que *podrían* suceder en una situación. Según la idea que estoy expresando, es simplemente *imposibilidad* de que pudieran formarse agujeros negros en alguna etapa (y en consecuencia, se destruya información) la que debe ser compensada por la falta de determinismo en la teoría cuántica.

¿CUÁNDO SE REDUCE EL VECTOR DE ESTADO?

Supóngase que aceptamos, sobre la base de los argumentos precedentes, que la reducción del vector de estado pudiera ser de alguna forma un fenómeno gravitatorio en última instancia. ¿Pueden hacerse más explícitas las relaciones entre R y la gravitación? Sobre la base de esta idea, ¿cuándo tendría lugar *realmente* el colapso del vector de estado?

Señalaré primero que incluso en las aproximaciones más "convencionales" a la teoría de la gravitación cuántica existen algunas dificultades técnicas serias para conciliar los principios de la relatividad general con las reglas de la teoría cuántica. Estas reglas (principalmente la forma de reinterpretar el momento como derivada respecto a la posición, en la expresión para la ecuación de Schrödinger) no se ajustan en absoluto a las ideas de la geometría del espacio-tiempo curvo. Mi punto de vista es que tan pronto como se introduce una cantidad "significativa" de curvatura espacio-temporal, las reglas de la superposición lineal cuántica deben fallar. Es en este punto donde las superposiciones de amplitudes complejas de estados potenciales alternativos quedan reemplazadas por opciones reales con pesos probabilísticos, y una de estas alternativas tiene lugar *realmente.*

¿Qué entiendo por una cantidad "significativa" de curvatura? Entiendo que se ha alcanzado el nivel en el que la medida de la curvatura introducida tiene aproximadamente la escala de *un gravitón*[6] *o* más. (Recuérdese que, según las reglas de la teoría cuántica, el campo electromagnético está "cuantizado" en unidades individuales, llamadas "fotones". Cuando se descompone el campo en sus frecuencias individuales, la parte de frecuencia v puede llegar sólo en números enteros de fotones, cada uno de energía *hv.* Reglas análogas se aplicarán presumiblemente al campo gravitatorio.) Un gravitón sería la unidad más pequeña de curvatura

[6] Debe estar permitido que existan los llamados *gravitones longitudinales:* los gravitones "virtuales" que componen un campo gravitatorio estático. Desgraciadamente hay problemas teóricos implicados al definir tales cosas de una forma matemática precisa e invariante".

que estaría permitida por la teoría cuántica. La idea es que, *en* cuanto se alcanza este nivel, las reglas ordinarias de superposición lineal, de acuerdo con el procedimiento U, se modificarán cuando se apliquen a gravitones, y se establecerá algún tipo de "inestabilidad no lineal" con asimetría temporal. En lugar de tener para siempre superposiciones lineales complejas de "opciones" coexistentes, una de las posibilidades vence en esta etapa y el sistema "cae" en una alternativa u otra. Quizá la elección de alternativa se hace por simple azar, o quizá hay algo más profundo que subyace a esta elección. Pero ahora, una u otra se ha hecho *realidad.* El procedimiento R ha tenido efecto.

Nótese que, según esta idea, el procedimiento R ocurre espontáneamente de una manera totalmente objetiva, independiente de cualquier intervención humana. La idea es que el nivel de "un gravitón" estaría cómodamente situado entre el "nivel cuántico" de átomos, moléculas, etc., en donde son válidas las reglas lineales (U) de la teoría cuántica ordinaria, y el "nivel clásico" de nuestras experiencias cotidianas. ¿Cuál *es* la "magnitud" de este nivel de un gravitón? Debe subrayarse que no se trata realmente de una cuestión de *tamaño* físico: es más una cuestión de distribución de masa y energía. Hemos visto que los efectos de interferencia cuántica pueden ocurrir a distancias muy grandes siempre que no haya mucha energía implicada. (Recuérdese la autointerferencia del fotón descrita, y los experimentos EPR de Clauser y Aspect.) La escala de masas característica de la gravitación cuántica es la que se conoce como la *masa de Planck*

$$m_p = 10^{-5} \text{ gramos}$$

(aproximadamente). Esto parece ser bastante mayor de lo que nos gustaría, ya que objetos de masa mucho *menor* que ésta, como motas de polvo, pueden percibirse directamente comportándose de forma clásica. (La masa m_p es un poco mas pequeña que la de una pulga).

Sin embargo, no creo que el criterio de un gravitón tuviera que aplicarse de un modo tan crudo como este. Trataré de ser un poco mas explícito, pero en el momento de escribir este libro sigue habiendo demasiadas oscuridad y ambigüedades como para que este criterio se aplique exactamente.

Hay una manera muy directa para observar una partícula: utilizando una *cámara de niebla* de Wilson. En este caso tenemos una cámara llena de vapor que está apenas por encima del punto de condensación en gotas. Cuando una partícula cargada que se mueve rápidamente entra en esta cámara, habiendo sido producida, pongamos por caso, por la desintegración de un átomo radioactivo situado fuera de la cámara, su paso a través del vapor provoca que algunos átomos próximos a su camino le ionicen (es decir, se carguen eléctricamente debido a que algunos electrones son desalojados). Estos átomos ionizados actúan como centros en los que el vapor se condensa en pequeñas gotas. De esta forma, tenemos una traza de gotas que el experimentador puede observar directamente (fig. VIII.7).

Ahora bien, ¿cuál es la descripción mecánico-cuántica de esto? En el instante en que se desintegra nuestro átomo radioactivo, emite una partícula. Pero existen muchas direcciones posibles en las que podría viajar dicha partícula. Habrá una amplitud para esta dirección, una amplitud para aquella dirección, y una amplitud para cualquier otra dirección, todas ellas ocurriendo simultáneamente en una superposición lineal cuántica. La totalidad de todas estas opciones superpuestas constituye una onda esférica que emana del átomo desintegrado: la *función de onda* de la partícula emitida. A medida que cada posible traza de la partícula entra en la cámara de niebla queda asociado a una cadena de átomos ionizados, cada uno de los cuales

comienza a actuar como un centro de condensación para el vapor. Todas estas diferentes cadenas posibles de átomos ionizados deben coexistir también en una superposición lineal cuántica, de modo que ahora tenemos una superposición lineal de un gran número de *diferentes* cadenas de gotas condensadas. En cierta etapa, esta superposición lineal cuántica compleja se transforma en una colección de posibilidades *reales* con pesos probabilistas también reales que, según el procedimiento R, son los cuadrados de los módulos de las amplitudes de probabilidad. Solamente *una* de estas posibilidades se realiza en el mundo físico real de la experiencia, y esta alternativa particular es la que observa el experimentador. Según el punto de vista que estoy proponiendo, esta etapa ocurre en cuanto la diferencia entre los campos gravitatorios de las diversas posibilidades *alcanza* el nivel de un gravitón.

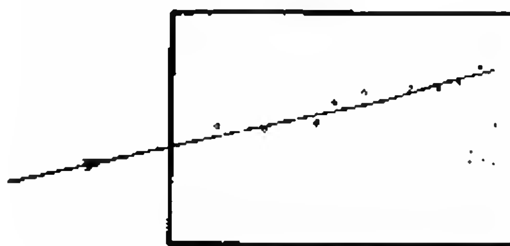


FIGURA VIII.7. *Una partícula cargada que entra en una cámara de niebla de Wilson y provoca la condensación de una cadena de gotas.*

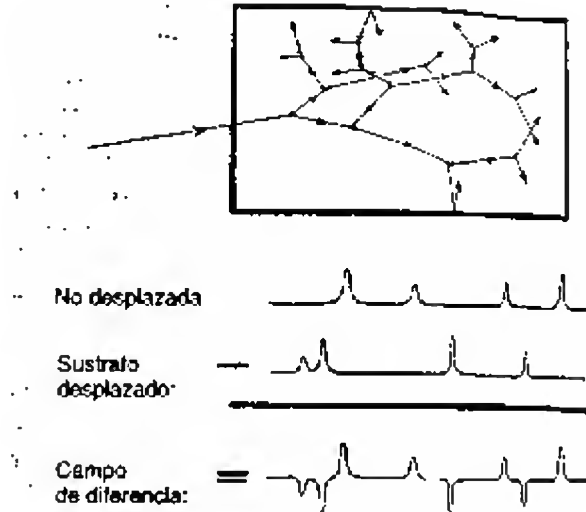
¿Cuándo sucede esto? Según un cálculo rápido,[7] si hubiera habido simplemente *una* gota esférica completamente uniforme, la etapa de un gravitón se alcanzaría cuando la gota crezca hasta aproximadamente una centésima de m_p, que es una diezmillonésima de gramo. Existen muchas incertidumbres en este cálculo (incluyendo algunas dificultades de principio) y el tamaño es un poco grande para que nos quedemos tranquilos, pero el resultado no es totalmente irrazonable. Hay que esperar obtener más adelante algunos resultados más precisos y será posible tratar una cadena entera de gotas y no solamente una simple gota. Puede que también haya importantes diferencias cuando se tenga en cuenta el hecho de que las gotas se componen de un número muy grande de átomos diminutos, en lugar de ser totalmente uniformes. Además, el propio criterio de "un gravitón" necesita hacerse mucho más preciso matemáticamente.

En la situación anterior he considerado que podría haber una observación real de un proceso cuántico (la desintegración de un átomo radioactivo) en el que los efectos cuánticos se han amplificado hasta el punto en que las diferentes opciones cuánticas producen diferentes, y directamente observables, opciones macroscópicas. Mi idea sería que R puede tener lugar objetivamente incluso cuando no está *manifiestamente* presente tal amplificación. Supongamos que, en lugar de entrar en una cámara de niebla, nuestra partícula entra simplemente en una gran caja de gas (o fluido) de tal densidad que es prácticamente seguro que chocará, o perturbará de alguna otra manera, con un gran número de átomos del gas. Consideremos sólo dos de las posibilidades para la partícula como parte de la superposición lineal compleja inicial: podría no entrar en absoluto en la caja, o podría entrar a lo largo de un camino particular y rebotar en algún átomo del gas. En el segundo caso, dicho átomo de gas saldrá despedido a gran velocidad de una forma en que no lo hubiera hecho si la partícula no hubiera incidido sobre él, y a continuación

[7] Mis propios burdos intentos originales para calcular este valor fueron mejorados por Abhay Ashtekar, y estoy utilizando aquí su valor (véase Penrose, 1987a). Sin embargo, él me ha enfatizado que hay bastante arbitrariedad en algunas de las hipótesis que parecemos obligados a utilizar, de modo que hay que tener mucha precaución al adoptar el valor exacto obtenido para la masa.

chocará y él mismo rebotará en algún otro átomo más. Cada uno de los dos átomos se moverán entonces de formas en que no lo hubieran hecho en otro caso, y pronto habrá una cascada de movimiento de átomos en el gas que no habría sucedido si la partícula no hubiera entrado inicialmente en la caja (fig. VIII.8).

En este segundo caso no transcurriría mucho tiempo hasta que prácticamente todos los átomos en el gas hubieran sido perturbados por este movimiento.



Campos gravitacionales (altamente esquemáticos) de partículas

FIGURA VIII.8. Si *una. partícula entra en una gran caja llena de algún gas no pasará mucho tiempo hasta que prácticamente todos los átomos del gas hayan sido perturbados. Una superposición lineal cuántica de la partícula entrante, y de la panícula no entrante, implicaría la superposición lineal de dos geometrías espacio-temporales diferentes que describen los campos gravitatorios de las dos configuraciones de las partículas del gas. ¿Cuándo se alcanzará el nivel de un gravitón para la* diferencia *entre estas geometrías?*

Pensemos ahora en cómo deberíamos describir esto mecánico-cuánticamente. Inicialmente sólo tenemos la partícula original cuyas diferentes posiciones debe considerarse que ocurren en una superposición lineal compleja, como parte de la función de onda de la partícula. Pero tras un corto intervalo, todos los átomos del gas deben estar involucrados. Consideremos la superposición lineal compleja de dos caminos que hubiera podido tomar la partícula, uno entrando en la caja y el otro no. La mecánica cuántica estándar insiste en que extendamos esta superposición a todos los átomos del gas: debemos superponer dos estados tales que todos los átomos de gas en uno de los estados estén desplazados respecto a sus posiciones en el otro estado. Consideremos ahora la *diferencia* entre los campos gravitatorios de la totalidad de los átomos individuales. Incluso aunque la distribución *global* del gas sea virtualmente la misma en los dos estados a superponer (y los campos gravitatorios globales pudieran ser prácticamente idénticos), si *restamos* un campo de otro obtenemos un campo *diferencia* (fuertemente oscilante) __que podría ser perfectamente "significativo" en el sentido en que lo entiendo aquí— y puede que se supere fácilmente el nivel de un gravitón para el campo *diferencia.* En cuanto se alcanza este nivel tiene lugar la reducción del vector de estado: en el estado *real* del sistema, *o bien* la partícula ha entrado en la caja de gas, *o bien* no lo ha hecho. La superposición lineal compleja se

ha reducido a posibilidades estadísticamente ponderadas, y solamente *una* de estas tiene lugar realmente.

En el ejemplo previo consideré una cámara de niebla como una manera de proporcionar una observación mecánico-cuántica. Considero muy probable que se puedan tratar otros tipos de observaciones semejantes (placas fotográficas, cámaras de chispas, etc.) utilizando el "criterio de un-gravitón" mediante una aproximación semejante a la que he indicado antes para la caja de gas. Queda mucho trabajo por hacer para ver cómo podría aplicarse en detalle este procedimiento.

Por ahora éste es sólo el germen de una idea para la que creo muy necesaria una nueva teoría.[8] Pienso que cualquier esquema completamente satisfactorio tendría que incorporar algunas ideas radicalmente nuevas acerca de la naturaleza de la geometría del espacio-tiempo, incluyendo probablemente una descripción esencialmente no local.[9] Una de las razones más imperativas para creer esto procede de los experimentos tipo EPR, en los que una "observación" (en este caso, el registro de una fotocélula) en un extremo de una habitación puede efectuar la reducción *simultánea* del vector de estado en el otro extremo. La construcción de una teoría completamente objetiva de la reducción del vector de estado que sea consistente con el espíritu de la relatividad es un profundo reto, ya que "simultaneidad" es un concepto extraño a la relatividad por ser dependiente del movimiento del observador. Mi opinión es que nuestra imagen actual de la realidad física, especialmente en relación con la naturaleza del *tiempo* va a padecer una gran conmoción incluso mayor, quizá, que la que ya han supuesto hasta ahora la relatividad y la mecánica cuántica.

Debemos volver a nuestra pregunta original. ¿Cómo se relaciona todo esto con la física que gobierna las acciones de nuestros cerebros? - ¿Que tendría que ver con nuestros pensamientos y nuestros sentimientos? para intentar algún tipo de respuesta será necesario en primer lugar examinar algo de cómo están construidos realmente nuestros cerebros. Volveré después a la que creo que es la pregunta fundamental: ¿qué tipo de acción *física* nueva está probablemente implícita cuando pensamos o percibimos conscientemente?

---

[8] Otros diversos intentos para proporcionar una teoría objetiva de la reducción del vector de estado han aparecido de tarde en tarde en la literatura. Los más importantes son Károlyházy (1974), Károlyházy, Frenkel y Lukács (1986), Komar (1969), Pearle (1985, 1989), Ghirardi, Rimini y Weber (1986).

[9] Yo mismo he estado interesado, durante años, en tratar de desarrollar una teoría no-local del espacio-tiempo, motivada generalmente desde otras direcciones, denominada "teoría de twistor" (véase Penrose y Rindler, 1986; Huggett y Tod, 1985; Ward y Wells, 1990). Sin embargo, a esta teoría aún le faltan, en el mejor de los casos, algunos ingredientes esenciales, y no sería apropiado entrar ahora en una discusión sobre ella.

IX. CEREBROS REALES Y MODELOS DE CEREBRO

¿CÓMO SON REALMENTE LOS CEREBROS?

EN EL INTERIOR DE NUESTRAS CABEZAS hay una magnífica estructura que controla nuestras acciones y de algún modo da lugar a una conciencia del mundo que nos rodea. Pero, como Alan Turing dijo en cierta ocasión,[1] ¡no hay nada más parecido a un puchero de potaje! Es difícil ver cómo un objeto de apariencia tan poco prometedora pueda lograr los milagros de que le sabemos *capaz.* Sin embargo, un examen más próximo comienza a revelar que el cerebro tiene una estructura mucho más intrincada y una sofisticada organización (fig. IX. 1). La parte superior, con más circunvoluciones (y la más parecida al potaje), es el *cerebro* propiamente dicho. Está claramente dividido por la mitad en los *hemisferios cerebrales* izquierdo y derecho; y, de una manera bastante menos tajante, en una zona delantera con el lóbulo frontal y una zona trasera con otros tres lóbulos: el parietal, el temporal y el occipital. Debajo y en la parte de atrás hay una porción bastante más pequeña y algo esférica (parecida quizá a dos ovillos de lana): el *cerebelo.* Más en el interior, y parcialmente ocultas bajo el cerebro, hay varias estructuras curiosas y de apariencia complicada: el puente y la médula (incluyendo la formación reticular, una región que nos interesará más adelante) que constituyen el tronco cerebral, tálamo, hipotálamo, hipocampo, cuerpo calloso, y muchas otras construcciones extrañas y de nombres singulares.

El cerebro[*] propiamente dicho es la parte de la que los seres humanos se sienten más orgullosos, pues no sólo es la parte más grande del cerebro humano sino que es también mayor, en proporción al encéfalo en conjunto, en el *hombre* que en los otros animales. (También el *cerebelo* es mayor en el hombre que en la mayoría de los animales.) El cerebro y el cerebelo tienen capas superficiales externas relativamente delgadas de *sustancia gris* y regiones internas mayores de *sustancia blanca.* Estas regiones de sustancia gris se denominan respectivamente *corteza cerebral* y *corteza cerebelar.* La sustancia gris es en donde parece que se ejecutan los diversos tipos de tareas computacionales, mientras que la sustancia blanca consiste en largas fibras nerviosas que transportan señales de una parte del cerebro a otra.

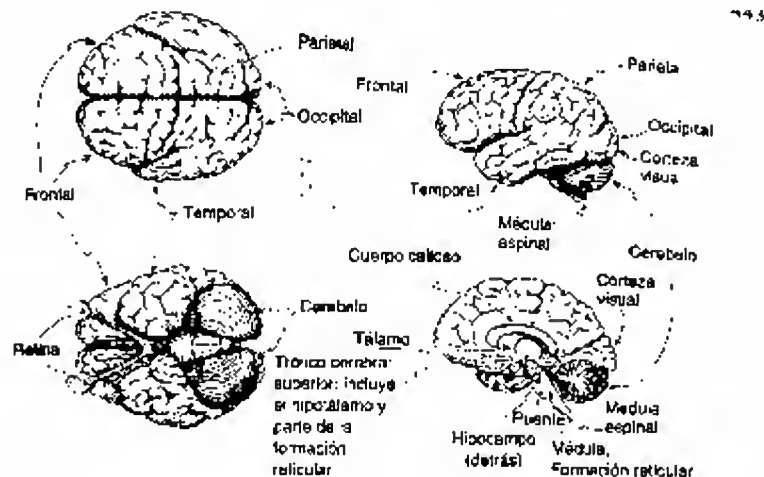


FIGURA IX.1 *El cerebro humano: vistas superior, lateral, inferior y sección central.*

---

[1] En una emisión radiofónica de la BBC; véase Hodges (1983), p. 419.

[*] A lo largo de este capítulo habrá que tener en cuenta la distinción más precisa entre cerebro global, o encéfalo, y cerebro propiamente dicho, que es una parte del anterior. [N. del T.]

Diversas partes de la corteza cerebral están asociadas con funciones específicas. La *corteza visual* es una región en el interior del lóbulo occipital, justo en la parte trasera del cerebro, que está relacionada con la recepción e interpretación de la visión. Es curioso que la naturaleza escogiera esta región para reinterpretar las señales procedentes de los ojos que, al menos en el hombre, están situados justo en *la parte frontal* de la cabeza. Pero la naturaleza se comporta de forma aún más curiosa que ésta. Es el hemisferio cerebral *derecho* el que está relacionado casi exclusivamente con el lado *izquierdo* del cuerpo, mientras que el hemisferio cerebral izquierdo está relacionado con el lado derecho del cuerpo, de modo que prácticamente todos los nervios deben cruzar de un lado a otro cuando entran o salen del cerebro. En el caso de la corteza visual, no se trata simplemente de que el lado derecho esté asociado al ojo izquierdo sino que está asociado con *el lado izquierdo del campo visual de ambos ojos.* Análogamente, la corteza visual izquierda está asociada con el lado derecho del campo visual de ambos ojos. Esto significa que los nervios que salen del lado derecho de la retina de cada ojo deben ir a la corteza visual derecha (recuérdese que la imagen en la retina está invertida) y que los nervios que salen del lado izquierdo de la retina de cada ojo deben ir a la corteza visual izquierda. (Véase fig. IX.2.) De este modo se forma un mapa muy bien definido del lado izquierdo del campo *visual* en la corteza visual derecha y se forma otro mapa del lado derecho del campo visual en la corteza visual izquierda.

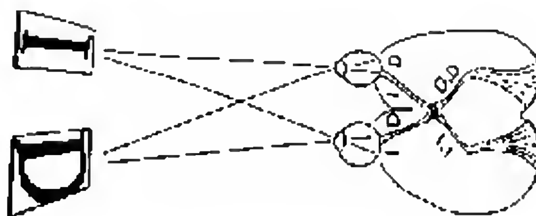


FIGURA IX.2. *El lado izquierdo del campo visual de ambos ojos se proyecta en la corteza visual derecha, y el lado derecho del campo visual se proyecta en la corteza visual izquierda. Vista inferior: nótese que las imágenes en la retina están invertidas.*

Las señales de los oídos también cruzan al lado contrario del cerebro de esta curiosa manera. La corteza auditiva derecha (parte del lóbulo temporal derecho) procesa principalmente el sonido recibido desde la izquierda, y la corteza auditiva izquierda, en general, los sonidos que proceden de la derecha. El olfato parece una excepción a las reglas generales. La corteza olfativa derecha, situada en la parte frontal del cerebro (en el lóbulo frontal —lo que es excepcional para un área sensorial—), está relacionada principalmente con la ventana derecha de la nariz y la izquierda, con la ventana izquierda.

Las sensaciones del *tacto* tienen que ver con la región del lóbulo parietal denominada *corteza somatosensorial.* Esta región está exactamente detrás de la división entre los lóbulos frontal y parietal. Existe una correspondencia muy concreta entre las diversas partes de la superficie del cuerpo y las regiones de la corteza somatosensorial. Esta correspondencia se ilustra a veces gráficamente en términos de lo que se denomina el "homúnculo somatosensorial", que es una figura humana distorsionada que se representa yaciendo a lo largo de la corteza somatosensorial como en la fig. IX.3. La corteza somatosensorial derecha trata las sensaciones del lado izquierdo del cuerpo, y la izquierda, las del lado derecho. Existe una región análoga en el lóbulo *frontal,* situada justo *delante* de la división entre el lóbulo frontal y el parietal, conocida como corteza *motora.* Ésta está relacionada con la activación del *movimiento* de las diferentes partes del

cuerpo y de nuevo existe una correspondencia muy específica entre los diversos músculos del cuerpo y las regiones de la *corteza* motora. Ahora tenemos un "homúnculo motor" para representar esta correspondencia, como se muestra en la fig. IX.4. La corteza motora derecha controla el lado izquierdo del cuerpo, y la corteza motora izquierda controla el lado derecho.

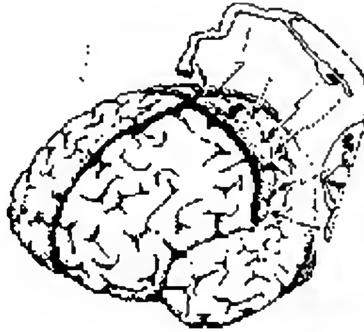


FIGURA IX 3. *El "homúnculo somatosensorial" ilustra gráficamente las porciones del cerebro —inmediatamente detrás de la división entre los lóbulos frontal y parietal— que están más directamente interesadas en el sentido del tacto para las diversas panes del cuerpo.*

Las regiones de la corteza cerebral recién mencionadas (las cortezas visual, auditiva, olfativa, somatosensorial y motora) se llaman *primarias* puesto que son las más directamente relacionadas con el *input* y el *output* del cerebro. Próximas a estas regiones primarias están las regiones *secundarias* de la corteza cerebral, que están relacionadas con un nivel de abstracción más sutil y complejo. (Véase fig. IX.5.) La información

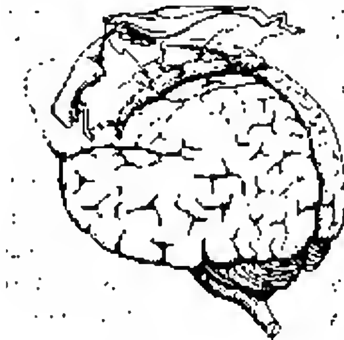


FIGURA IX.4. *El "homúnculo motor" ilustra las porciones del cerebro —inmediatamente delante de la división entre los lóbulos frontal y parietal— que más directamente activan los movimientos de las diversas partes del cuerpo.*

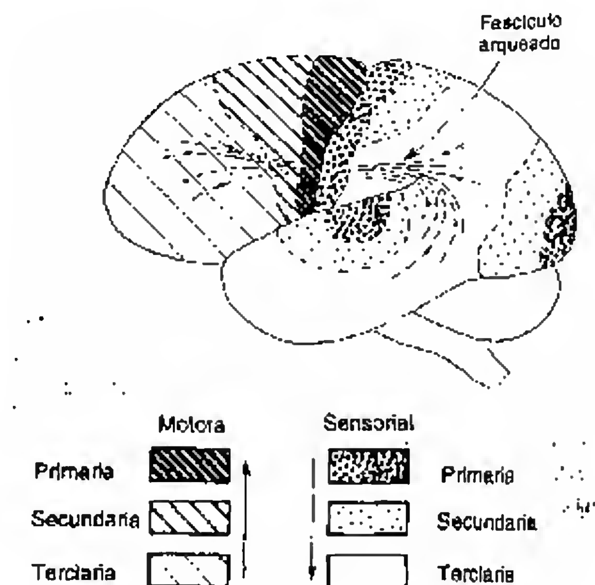


FIGURA IX.5. *La acción del cerebro en líneas muy generales. Los datos externos de los sentidos entran en las regiones sensoriales primarias, son procesados con sucesivos grados de sofisticación en las regiones sensoriales secundarias y tercianas, transferidos a la región motora terciaria y, finalmente, son transformados en instrucciones específicas para el movimiento en las regiones motoras primarias.*

sensorial recibida por las cortezas visual, auditiva y somatosensorial se procesa en las regiones secundarias asociadas, y la región motora secundaria está relacionada con los planes concebidos del movimiento que son traducidos por la corteza motora primaria en instrucciones más específicas para los movimientos musculares reales. (Dejemos aparte la corteza olfativa en nuestras consideraciones, ya que se comporta de modo diferente y parece que se conoce más bien poco sobre ella.) Las restantes regiones de la corteza cerebral se denominan *tercianas* (o cortezas *de asociación).* Es en estas regiones terciarias en donde se lleva a cabo fundamentalmente la actividad más abstracta y sofisticada del cerebro. Aquí es —en conjunción, en alguna medida, con la periferia— donde se mezcla y analiza la información de las diversas regiones sensoriales de una manera muy compleja, donde reside la memoria, se construyen las imágenes del mundo externo, se conciben y evalúan planes generales, y se entiende o se formula el habla.

El habla es particularmente interesante ya que se suele considerar como algo muy específico de la inteligencia humana. Es curioso que (al menos en la inmensa mayoría de las personas diestras y en la mayor parte de las personas zurdas) los principales centros del habla están precisamente en el *lado izquierdo* del cerebro. Las áreas esenciales son el *área de Broca,* una región en la parte inferior trasera del lóbulo frontal y otra llamada *área de Wernicke,* dentro y también alrededor de la parte superior trasera del lóbulo temporal (véase fig. IX.6). El área de Broca está relacionada con la formulación de enunciados, y el área de Wernicke con la comprensión del lenguaje. Las lesiones en el área de Broca dificultan el habla pero dejan intacta la comprensión, mientras que con el área de Wernicke lesionada el habla es fluida pero con poco contenido. Un haz nervioso llamado *fascículo arqueado* conecta las dos áreas. Cuando éste se lesiona no queda impedida la comprensión y el habla sigue siendo fluida, pero la comprensión no puede ser expresada verbalmente. Podemos formarnos ahora una imagen muy general de lo que hace el cerebro. El *input* del cerebro procede de las señales visuales, auditivas, táctiles y otras que se registran

inicialmente en las porciones *primarias de* (principalmente) los lóbulos *posteriores* (parietal, temporal y occipital). El *output* del cerebro, en forma de activación de movimientos corporales, se produce principalmente en las porciones primarias de los lóbulos *frontales* del cerebro. Entre ambos tiene lugar algún tipo de procesamiento. De un modo general, existe un movimiento en la actividad cerebral que comienza en las porciones primarias de los lóbulos posteriores, se desplaza a las porciones secundarias a medida que se van analizando los datos de entrada, y continúa hacia las porciones terciarias de los lóbulos posteriores a medida que estos datos se van comprendiendo completamente *(v.g.* como sucede con la comprensión del habla en el área de Wernicke). El fascículo arqueado —el haz de fibras nerviosas antes mencionado, pero ahora en ambos lados del cerebro— lleva entonces esta información procesada al lóbulo frontal, en cuyas regiones terciarias se formulan planes generales de actuación *(v.g.* como la formulación del habla en el área de Broca).

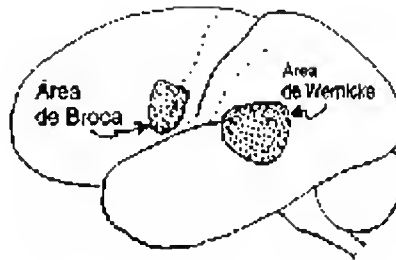


FIGURA IX.6. *Normalmente sólo en el lado izquierdo: el área de Wernicke está relacionada con la comprensión y el área de Broca lo está con la formulación del habla.*

Estos planes generales de actuación se traducen en concepciones más concretas sobre movimientos corporales en las regiones motoras secundarias y, finalmente, la actividad cerebral se mueve hacia la corteza motora primaria desde donde finalmente se envían las señales a los diversos grupos de músculos en el cuerpo (y a menudo a varios a la vez).

La imagen de un soberbio dispositivo de computación parece presentarse ante nosotros. Los defensores de la IA fuerte *(cfr.* capítulo I) sostendrán que aquí tenemos un ejemplo supremo de una computadora algorítmica —una máquina de Turing en efecto— en donde hay un *input* (como la cinta *input a* la izquierda de una máquina de Turing) y un *output* (como la cinta *output a* la derecha de la máquina), y entre las dos a mitades se realizan todo tipo de computaciones complicadas. Por supuesto, la actividad del cerebro puede llevarse a cabo también independientemente de cualquier input sensorial. Esto sucede cuando simplemente pensamos, calculamos o meditamos sobre recuerdos. Para los defensores de la IA fuerte estas actividades del cerebro serían simplemente actividad algorítmica adicional, y ellos podrían sugerir que el fenómeno de la "conciencia" aparece allí donde semejante actividad interna alcanza un nivel de sofisticación.

Sin embargo, no debemos conformarnos con estas rápidas explicaciones. La imagen general de la actividad del cerebro presentada arriba es solamente una imagen bastante tosca. En primer lugar, incluso la recepción de la visión no es tan sencilla como la he presentado. Parece haber varias regiones diferentes (aunque más pequeñas) de la corteza en donde se hacen mapas del campo visual, aparentemente con otros diversos propósitos. (Nuestra *conciencia* de la visión parece diferir con respecto a ellos.) Parece que hay también otras regiones sensoriales y motoras diseminadas por la corteza cerebral (por ejemplo, los movimientos del ojo pueden ser activados por varios puntos en los lóbulos *posteriores).*

En mis descripciones anteriores no he mencionado siquiera el papel de las partes del encéfalo distintas del cerebro propiamente dicho. ¿Cuál es, por ejemplo, el papel del *cerebelo?* Aparentemente es el responsable de una precisa coordinación y control del cuerpo, su ritmo, equilibrio y delicadeza de movimientos. Imaginemos el arte fluido de un bailarín, la fácil precisión de un jugador de tenis profesional, el rapidísimo control de un piloto de carreras, y los movimientos seguros de las manos de un pintor o un intérprete musical; imaginemos también los gráciles saltos de una gacela y el andar sigiloso de un gato. Tal precisión no sería posible sin el cerebelo, y todos los movimientos se volverían titubeantes y torpes. Parece que cuando se está aprendiendo una nueva habilidad, ya sea andar o conducir un coche, se debe planear inicialmente cada acción en detalle y es el cerebro quien controla; pero cuando ya ha sido dominada esta habilidad —y se ha convertido en una "segunda naturaleza"—, es el cerebelo el que asume el mando. Además, resulta una experiencia familiar que si *pensamos* nuestras acciones en una habilidad que ya se domina, entonces podemos perder momentáneamente el control. La *reflexión* parece implicar la reintroducción del control cerebral y, aunque así se introduce una consiguiente flexibilidad en la actividad, se pierde la acción cerebelar precisa y fluida. Tales descripciones están sin duda muy simplificadas pero dan una impresión razonable del papel del cerebelo.*

Resultaba también confuso, en mis anteriores descripciones de la acción del cerebro, el dejar de lado cualquier información sobre las demás partes del encéfalo. Por ejemplo, el *hipocampo* juega un papel vital al asentar la memoria a largo plazo (permanente), siendo almacenada la memoria real en alguna parte de la corteza cerebral —probablemente en varios lugares a la vez. El cerebro puede retener también imágenes a *corto* plazo de otras maneras; y puede mantenerlas durante algunos minutos o incluso horas (quizá "conservándolas en la mente"). Pero para poder recordar estas imágenes después de que hayamos dejado de prestarles atención es necesario que queden asentadas de forma permanente, y para esto es esencial el hipocampo. (Las lesiones en el hipocampo provocan una temible condición en la que no se retienen nuevos recuerdos una vez que han dejado la atención del sujeto.) El *cuerpo calloso* es la región a través de la que se comunican entre sí los hemisferios cerebrales izquierdo y derecho. (Veremos más adelante algunas de las sorprendentes consecuencias de tener el cuerpo calloso seccionado.) El *hipotálamo* es la sede de las emociones —placer, rabia, miedo, desesperación, hambre— y sirve de medio a las manifestaciones tanto mentales como físicas de las emociones. Hay un flujo continuo de señales entre el hipotálamo y las diferentes partes del cerebro. El *tálamo* actúa como un importante centro de procesado y estación repetidora, y transmite muchos de los impulsos nerviosos desde el mundo externo a la corteza cerebral. La *formación reticular* es responsable del estado general de alerta o conciencia implicado en el encéfalo como un todo o en las distintas partes del encéfalo. Hay numerosos caminos para los nervios que conectan estas, y muchas otras, áreas de importancia vital.

La descripción anterior es sólo una muestra de las partes más importantes del encéfalo. Concluiré la sección diciendo algo más sobre la organización del encéfalo en general. Sus diferentes partes se clasifican en tres regiones que, tomadas en orden según se alejan de la médula espinal, se llaman *cerebro posterior* (o rombencéfalo), el *cerebro medio* (o mesencéfalo) y el *cerebro anterior* (o prosencéfalo).

* Curiosamente, el comportamiento "cruzado" del cerebro no se aplica al cerebelo, de modo que la mitad derecha del cerebelo controla principalmente el lado *derecho* del cuerpo, y la mitad izquierda controla el lado *izquierdo* del cuerpo.

En el desarrollo temprano del embrión se encuentran estas tres regiones en este orden, como tres abultamientos al final de la médula espinal De la que está situada en la parte más extrema, el cerebro anterior en fase de desarrollo, brotan dos yemas, una a cada lado, que se convertirán en los hemisferios cerebrales. El cerebro anterior completamente desarrollado incluye muchas partes importantes del encéfalo, no sólo el cerebro propiamente dicho, sino también el cuerpo calloso, tálamo, hipotálamo, hipocampo y muchas otras partes. El cerebelo es parte del cerebro posterior. La formación reticular tiene una parte en el cerebro medio y otra parte en el cerebro posterior. El cerebro anterior es el "más reciente" en el sentido del desarrollo evolutivo, y el cerebro posterior es el más "antiguo".

Espero que este breve cuadro, aunque impreciso en diversos aspectos, dará al lector alguna idea de cómo es el cerebro humano y qué hace de manera general. Hasta aquí apenas he tocado el tema central de la *conciencia.* Abordaremos este tema a continuación.

¿DÓNDE ESTÁ LA SEDE DE LA CONCIENCIA?.

Se han expresado muchas opiniones diferentes con respecto a la relación entre el estado del cerebro y el fenómeno de la conciencia. Es notable el escaso consenso de opinión para un fenómeno de tan obvia importancia. Es evidente, sin embargo, que no todas las partes del cerebro están involucradas por igual en su manifestación. Por ejemplo, como se apuntó antes, el cerebelo parece ser mucho más "autómata" que el cerebro propiamente dicho. Las acciones bajo control cerebelar parecen tener lugar casi de forma "autónoma" sin que tengamos que "reflexionar" sobre ellas. Mientras que podemos decidir conscientemente el andar de un lugar a otro, no tenemos conciencia a menudo de los elaborados planes de movimientos musculares detallados que serán necesarios para el movimiento controlado. Lo mismo puede decirse de las acciones reflejas inconscientes, como la de retirar la mano de una estufa caliente, que podría estar mediada no por el cerebro en general sino por la parte superior de la médula espinal. A partir de esto, podemos estar inclinados, al menos, a inferir que es probable que el fenómeno de la conciencia tenga más que ver con la acción del cerebro propiamente dicho que con la del cerebelo o la médula espinal.

Por otra parte, está lejos de ser evidente que la actividad del cerebro deba *siempre* incidir sobre nuestra conciencia. Por ejemplo, como he descrito antes, en la acción normal de caminar en la que no somos conscientes de la actividad detallada de nuestros músculos y miembros —al ser el control de esta actividad principalmente cerebelar (asistido por otras partes del cerebro y la médula espinal)— parece que *también* las regiones motoras primarias deberían estar involucradas. Además lo mismo sería válido para las regiones sensoriales primarias: podríamos no tener conciencia, en ese momento, de las variaciones de presión en las plantas de nuestros pies cuando caminamos, pero las regiones correspondientes de nuestra corteza somatosensorial estarían siendo activadas continuamente.

De hecho, el distinguido neurocirujano canadiense-estadounidense Wilder Penfield (quien, en los años 40 y 50 fue responsable de detallar gran parte del mapa de las regiones sensorial y motora del cerebro humano) ha argumentado que la conciencia *no* está simplemente asociada a la actividad cerebral. Él sugirió, basado en sus experiencias al realizar numerosas operaciones cerebrales en sujetos concientes, que cierta región a la que denominó *tronco cerebral superior,* consistente principalmente en el tálamo y el cerebro medio *(cfr.* Penfield y Jasper, 1947) — aunque él tenía en mente sobre todo la formación reticular — debería considerarse, en cierto

sentido, como la "sede de la conciencia". El tronco cerebral superior está en comunicación con el cerebro, y Penfield argumentaba que la "atención consciente" o "conciencia de acción voluntaria" aparecería siempre que esta región del tronco cerebral estuviera en comunicación directa con la región apropiada de la corteza cerebral, es decir, la región particular asociada con cualesquiera sensaciones específicas, pensamientos, recuerdos o acciones que sean percibidas o evocadas conscientemente en el momento. Subrayó que aunque él pudiera, por ejemplo, estimular la región de la corteza motora del sujeto que provoca el movimiento del brazo derecho (y el brazo derecho se moviera realmente), esto no provocaría que el sujeto *quisiera* mover el brazo derecho. (De hecho, el sujeto podría incluso reaccionar con el brazo izquierdo y detener el movimiento del brazo derecho — como en la bien conocida interpretación cinematográfica que Peter Sellers hacía del doctor Strangelove.) * Penfield sugería que el *deseo* del movimiento podría tener más que ver con el tálamo que con la corteza cerebral. Su idea era que la conciencia es una manifestación de la actividad del tronco cerebral superior pero, puesto que se necesita además que haya algo que sea consciente de, no es sólo el tronco cerebral el que está implicado sino también alguna región de la corteza cerebral que esté en ese momento en comunicación con el tronco cerebral superior y cuya actividad representa el sujeto (impresión sensorial o recuerdo) o el objeto (acción voluntaria) de dicha conciencia.

Otros neurofisiólogos han argumentado que la formación reticular, en concreto, podría considerarse la "sede" de la conciencia, si realmente existe tal sede. Después de todo, la formación reticular es responsable del estado general de alerta del cerebro (Moruzzi y Magoun, 1949). Si se lesiona, el resultado es la inconciencia. Siempre que el cerebro está en un estado consciente de vigilia, la formación reticular está activa; de lo contrario no lo está. Parece haber así una clara asociación entre la actividad de la formación reticular y el estado de una persona que normalmente denominamos "consciente". Sin embargo, la cuestión se complica por el hecho de que en el estado de ensueño, en el que realmente se tiene "conciencia" (en el sentido de tener conciencia del propio sueño), las partes normalmente activas de la formación reticular *no* parecen estar activas. Una cosa que también preocupa a algunas personas, a propósito de asignar semejante honroso *status* a la formación reticular, es que ésta es, en términos evolutivos, una parte muy *antigua* del cerebro. Si todo lo que se necesita para ser consciente es una formación reticular activa, entonces las ranas, los lagartos e incluso los bacalaos son conscientes.

Personalmente no considero que este último argumento tenga mucha fuerza. ¿Qué evidencia tenemos de que los lagartos o los bacalaos *no* posean alguna forma de conciencia de bajo nivel? ¿Qué derecho tenemos a afirmar, como harían algunos, que los seres humanos son los únicos habitantes de nuestro planeta dotados de una capacidad real de tener "conciencia"? ¿Somos las únicas cosas, entre las criaturas de la Tierra, para quienes es posible "ser"? Lo dudo. Aunque las ranas y los lagartos, y especialmente los bacalaos, no me inspiran una gran convicción *de* que necesariamente hay "alguien ahí" devolviéndome la mirada cuando los observo, me resulta muy fuerte la impresión de "presencia consciente" cuando miro a un perro o un gato o, especialmente, cuando en el zoológico me mira un simio o un mono. No pido que ellos sientan como yo, ni siquiera que haya mucha sofisticación en su modo de sentir. No pido que sean "autoconscientes" en ningún sentido fuerte (aunque yo admitiría que puede estar presente un elemento de autoconciencia).* Todo lo que pido simplemente es que a veces *sientan*. Como en el estado de

* La película, dirigida por Stanley Kubrick, se tituló en México, *Dr. Insólito.* [N. del T.]

* Hay alguna evidencia convincente de que al menos los chimpancés son capaces de autoconciencia, como parecen demostrar los experimentos en los que se permite a los chimpancés que jueguen con espejos; *cfr.* Oakley (1985) capítulos IV y V.

ensueño, aceptaría que hay presente alguna forma de conciencia, aunque presumiblemente de un nivel muy bajo. Si partes de la formación reticular son, en cierto sentido, las únicas responsables de la conciencia, entonces deberían estar activas, aunque en un nivel bajo, en el estado de ensueño.

Otro punto de vista (O'Keefe, 1985) afirma que es la acción del *hipocampo* la que más tiene que ver con el estado consciente. Como señalé antes, el hipocampo es crucial para el asentamiento de la memoria a largo plazo. Podría alegarse que el asentamiento de recuerdos permanentes está asociado con la conciencia y, si esto es correcto, el hipocampo jugaría realmente un papel central en el fenómeno de la conciencia.

Otros sostendrán que es la propia corteza cerebral la responsable de la conciencia. Puesto que el cerebro propiamente dicho es el orgullo del hombre (aunque los cerebros de los delfines son tan grandes como el de éste) y puesto que las actividades mentales más estrechamente asociadas con la inteligencia parecen ser llevadas a cabo por este cerebro, entonces ¡es ciertamente allí donde reside el alma del hombre! Esta será presumiblemente la conclusión del punto de vista de la IA fuerte, por ejemplo. Si la "conciencia" es simplemente una característica de la *complejidad* de un algoritmo —o quizá de su "profundidad" o cierto "nivel de sutileza"— entonces, según la idea de la IA fuerte, los algoritmos complicados que ejecuta la corteza cerebral confirmarían a esta región como la más firme candidata a la capacidad de manifestar conciencia.

Muchos filósofos y psicólogos parecen aceptar la idea de que la conciencia humana está muy ligada al *lenguaje* humano. Por consiguiente, es sólo en virtud de nuestras capacidades lingüísticas por lo que podemos alcanzar una sutileza de pensamiento, que es la impronta misma de nuestra humanidad, y la expresión de nuestras propias almas. Es el lenguaje, según este punto de vista, el que nos distingue de los otros animales, y nos proporciona así una excusa para privarles de su libertad y sacrificarlos cuando sentimos que surge dicha necesidad. Es el lenguaje el que nos permite filosofar y describir cómo sentimos, de modo que podamos convencer a los demás de que *nosotros* tenemos conciencia del mundo exterior y también tenemos conciencia de nosotros mismos. Desde este punto de vista, nuestro lenguaje se considera como el ingrediente clave de nuestra posesión de conciencia.

Ahora bien, debemos recordar que nuestros centros del lenguaje están (en la inmensa mayoría de las personas) solamente en los lados *izquierdos* de nuestros cerebros (áreas de Broca y de Wernicke). El punto de vista recién expresado parecería implicar que la conciencia es algo que está asociado solamente con la corteza cerebral izquierda y no con la derecha. De hecho, ésta resulta ser la opinión de varios neurofisiólogos (en particular, John Eccles, 1973) aunque para mí, como profano, me parece realmente una idea muy extraña por las razones que voy a explicar.

EXPERIMENTO DE ESCISION CEREBRAL

Hay un importante conjunto de observaciones concernientes a humanos (y animales) a los que se ha seccionado completamente el cuerpo calloso, de modo que los hemisferios izquierdo y derecho de sus cortezas cerebrales no tienen comunicación entre sí. En el caso de los humanos,[2] el seccionamiento del cuerpo calloso fue realizado como operación terapéutica al descubrirse que

[2] Los primeros experimentos de este tipo se realizaron con gatos (*cfr.* Myers y Sperry, 1953). Para más información sobre experimentos de escisión cerebral, véase Sperry (1966), Gazzaniga (1970), MacKay (1987).

éste era un tratamiento efectivo para una forma de epilepsia particularmente grave que sufrían algunos sujetos. Numerosos *tests* psicológicos les fueron suministrados por Roger Sperry y sus colaboradores algún tiempo después de que hubieran sufrido estas operaciones. Los sujetos estaban colocados de tal forma que se les presentaban estímulos completamente separados para los campos izquierdo y derecho de visión, de modo que el hemisferio izquierdo sólo recibía información visual de lo que se mostraba en el lado derecho, y el hemisferio derecho, sólo de lo del lado izquierdo. Si se le mostraba brevemente una foto de un lápiz en el lado derecho y una foto de una copa en el izquierdo, el sujeto diría: "Eso es un *lápiz",* ya que era el lápiz, y no la copa, lo percibido por el único lado del cerebro aparentemente capaz de hablar. Sin embargo, el lado izquierdo sería capaz de seleccionar un plato, antes que un trozo de papel, como el objeto apropiado para asociar con la copa. El lado izquierdo estaría bajo el control del hemisferio derecho y, aunque incapaz de hablar, este hemisferio derecho ejecutaría ciertas acciones bastante complejas y decididamente humanas. De hecho, se ha sugerido que el *pensamiento geométrico* (especialmente en tres dimensiones), y también la música, pueden ser llevados a cabo normalmente en el hemisferio *derecho* principalmente, para compensar las capacidades verbales y analíticas del izquierdo. El cerebro derecho puede comprender los nombres comunes o las frases elementales, y puede llevar a cabo operaciones aritméticas muy sencillas.

Lo más sorprendente de estos sujetos con cerebro escindido es que los dos lados parecen comportarse como individuos prácticamente independientes, cada uno de los cuales puede comunicarse por separado con el experimentador —aunque la comunicación es más difícil—, y en un nivel más primitivo, con el hemisferio derecho que con el izquierdo, debido a la falta de capacidad verbal del derecho. Una mitad del cerebro del sujeto puede comunicar con la otra de manera simple, *v.g.* observando el movimiento del brazo controlado por el otro lado, o quizá oyendo sonidos reveladores (como el repiqueteo en un plato). Pero incluso esta comunicación primitiva entre los dos lados puede eliminarse en condiciones de laboratorio cuidadosamente controladas. Pese a todo todavía pueden pasar de un lado a otro vagos sentimientos emocionales debido presumiblemente, a que estructuras que no están seccionadas tales como el hipotálamo, siguen estando en comunicación con ambos lados.

Resulta tentador plantear la cuestión: ¿tenemos dos individuos conscientes separados que habitan en el mismo cuerpo? Esta cuestión ha sido objeto de mucha controversia. Algunos mantendrán que la respuesta debe ser ciertamente "sí", mientras otros afirmarán que ningún lado debe considerarse como un individuo por propio derecho. Algunos estarán de acuerdo en que el hecho de que puede haber sentimientos emocionales comunes a los dos lados pone en evidencia que todavía hay un solo individuo involucrado. Pero otro punto de vista es que sólo el hemisferio del lado *izquierdo* representa un individuo consciente, y el del lado derecho es un autómata. Parece que esta idea es la que aceptan quienes sostienen que el lenguaje es un ingrediente esencial de la conciencia. De hecho, sólo el hemisferio izquierdo puede responder convincentemente "Si" a la pregunta oral "¿Es usted consciente?" El hemisferio derecho, como un perro, un gato o un chimpancé, podría ser puesto en dificultades incluso para descifrar las palabras que constituyen la pregunta, y puede ser completamente incapaz de expresar verbalmente su respuesta.

Pero el tema no puede despacharse tan a la ligera. En un experimento más reciente de gran interés, Donald Wilson y sus colaboradores (Wilson *et al,* 1977; Gazzaniga, LeDoux y Wilson, 1977) examinaron un sujeto con cerebro escindido, a quien denominaban "PS" Después de la operación de seccionamiento sólo el hemisferio izquierdo podía hablar pero *ambos* hemisferios podían comprender el habla; posteriormente, el hemisferio derecho ¡aprendió también a hablar!

Evidentemente *ambos* hemisferios eran conscientes. Además, parecían ser conscientes *por separado,* ya que tenían diferentes gustos y deseos. Por ejemplo, el hemisferio izquierdo describía que su deseo era ser un dibujante y el derecho, ¡un piloto de carreras!

Por mi parte, simplemente no puedo creer la afirmación común de que el lenguaje humano ordinario es necesario para el pensamiento o para la conciencia. (Trataré de dar algunas de mis razones en el próximo capítulo.) Por lo tanto, me pongo de lado de quienes creen, en general, que las dos mitades de un sujeto con cerebro escindido pueden ser conscientes independientemente. El ejemplo de PS sugiere con fuerza que, al menos en este caso particular, ambas mitades pueden serlo realmente. En mi propia opinión la única diferencia real, a este respecto, entre PS y los demás es que su conciencia del lado derecho puede convencer realmente a los demás de su existencia.

Si aceptamos que PS tiene realmente dos mentes independientes, entonces se nos presenta una curiosa situación. Presumiblemente *antes* de la operación cada sujeto con cerebro escindido poseía una sola conciencia; pero después hay dos. De algún modo, la simple conciencia original se ha *bifurcado.* Podemos recordar al hipotético viajero del capítulo I que se sometía voluntariamente a la máquina teleportadora, y que (de forma imprevista) se despertaba para descubrir que su yo supuestamente "real" había llegado a Venus. En aquel caso, la bifurcación de su conciencia parecía proporcionar una paradoja. En efecto, podemos preguntar: "¿Qué ruta siguió 'realmente' su corriente de conciencia?" Si *usted* fuera el viajero, ¿cuál de ellos terminaría siendo "usted"? La máquina teleportadora podía despacharse como cosa de ciencia ficción, pero en el caso de PS parece que tenemos algo aparentemente análogo pero ¿qué *ha sucedido realmente?* ¿Cuál de las conciencias de PS "es" el PS de antes de la operación? Sin duda muchos filósofos desdeñarían la pregunta como carente de significado. Pero no parece haber modo operacional de decidir la cuestión. Cada hemisferio compartirá recuerdos de una existencia consciente antes de la operación y, sin duda, ambos afirmarán ser esa persona. Esto puede ser curioso pero no es en sí mismo una paradoja. De todas formas, aún nos sigue quedando algo de enigmático.

El enigma sería aún más exacerbado si las dos conciencias pudieran de algún modo ser unidas de nuevo con posterioridad. Volver a conectar los nervios individuales seccionados del cuerpo calloso parece imposible con la tecnología actual, pero podríamos concebir algo menos drástico que el seccionamiento real de las fibras nerviosas como situación inicial. Quizá estas fibras podrían ser congeladas temporalmente, o paralizadas con alguna droga. No tengo noticia de que se haya llevado a cabo un experimento semejante, pero supongo que podría ser técnicamente factible en un plazo no muy largo. Presumiblemente, después de que el cuerpo calloso se haya reactivado sólo *una* conciencia resultaría. Imagínese que esta conciencia es usted. ¿Cómo se sentiría después de haber sido dos personas separadas con distintos "yo" en algún momento en el pasado?

CEGUERA CORTICAL

Los experimentos de escisión cerebral parecen indicar al menos que no debe haber necesariamente una *única* "sede de la conciencia". Pero hay otros experimentos que parecen sugerir que algunas partes de la corteza cerebral están más asociadas a la conciencia que otras. Uno de estos tiene que ver con el fenómeno de la *ceguera cortical.* Lesiones en una región *de la*

corteza visual pueden provocar ceguera en el campo de visión correspondiente. Si se mantiene un objeto en dicha región del campo visual, entonces el objeto no será percibido. Se tiene ceguera con respecto a esa región de visión. Sin embargo, algunos curiosos descubrimientos *(cfr.* Weiskrantz 1987) indican que las cosas no son tan simples como eso. A un paciente de nominado "DB" hubo que extirparle parte de su corteza visual y esto le provocó la incapacidad para ver en una cierta región del campo visual Sin embargo, cuando se colocaba algo en esta región y se le pedía a DB que *adivinara* qué cosa era (normalmente una marca como una cruz o un círculo, o un segmento rectilíneo inclinado un cierto ángulo), él descubrió que podía hacerlo con casi 100% de exactitud. La precisión de estas "conjeturas" resultó una sorpresa para el propio DB, que seguía manteniendo que no podía percibir nada en dicha región.

Las imágenes recibidas en la retina son procesadas también en regiones del cerebro *distintas* de la corteza visual, y una de las regiones más oscuras involucradas está en la parte inferior del lóbulo temporal. Parece que DB podía estar basando sus "conjeturas" en información obtenida por esta región temporal inferior. Nada era directamente percibido *conscientemente* mediante la activación de estas regiones, pero la información estaba allí, para ser revelada sólo en la exactitud de las "conjeturas" de DB. De hecho, después de algún entrenamiento, DB fue capaz de obtener una cantidad limitada de conciencia real respecto a estas regiones.

Todo esto parece probar que algunas áreas de la corteza cerebral (v.g. la corteza visual) están más asociadas con la percepción consciente que otras áreas, pero que, con entrenamiento, algunas de estas otras áreas pueden aparentemente estar al alcance de la conciencia directa.

PROCESAMIENTO DE INFORMACIÓN EN LA CORTEZA VISUAL

La *corteza visual* es, más que cualquier otra parte del cerebro, la región de la que mejor se conoce cómo trata la información que recibe, y se han propuesto varios modelos para dar cuenta de esta acción. [3] En realidad, algún procesamiento de información visual tiene lugar en la misma [*] retina *antes de* que se alcance la corteza visual. (La retina se considera realmente como una parte del cerebro.) Uno de los primeros experimentos que señalaba cómo se lleva a cabo el procesamiento en la corteza visual les valió el Premio Nobel de 1981 a David Hubel y Torsten Wiesel. Estos científicos fueron capaces de demostrar que ciertas células de la corteza visual del gato respondían a líneas en el campo visual que tenían un *ángulo de inclinación particular.* Otras células próximas respondían a líneas con diferentes ángulos de inclinación. A menudo no importaba qué era lo que formaba este ángulo. Podía ser una línea que señalaba la frontera entre oscuridad y luz o, por el contrario, entre luz y oscuridad, o simplemente una línea oscura sobre un fondo claro. Sólo la característica del "ángulo de inclinación" había sido abstraída por las células concretas que se estaban examinando. Pero otras células respondían a colores concretos, o a las diferencias entre lo que percibe cada ojo de modo que se pueda obtener la percepción de profundidad. A medida que nos alejamos de las regiones de recepción primaria encontramos células que son sensibles a aspectos cada vez más sutiles de nuestra percepción de lo que vemos. Por ejemplo, la imagen de un triángulo blanco completo es percibida cuando miramos el dibujo de la fig. IX. 7; pero las líneas que forman el propio triángulo no están en realidad presentes en la

[3] Para un informe fácil de leer sobre el funcionamiento de la corteza visual, véase Hubel (1988).

[*] Algo en cierto modo complementario de la ceguera cortical es una situación conocida como "ceguera negada", según la cual un sujeto que de hecho es totalmente ciego insiste en que es *capaz* de ver bastante bien, ¡dando la impresión de ser *visualmente* consciente de los entornos *inferidos!* (Véase Churchland, 1984, p. 143.)

figura sino que son *inferidas.* Se han encontrado, en efecto, células en la corteza visual (en lo que se llama corteza visual secundaria) que pueden registrar las posiciones de estas líneas inferidas.

Hubo diversas afirmaciones en la bibliografía,[4] a comienzos de los años setenta, del descubrimiento de una célula en la corteza visual del mono que respondía sólo cuando se registraba en la retina la imagen de un rostro. Basada en esta información se formuló la "hipótesis de la célula de la abuela", según la cual habría ciertas células en el cerebro que responderían sólo cuando la abuela del sujeto entraba en la habitación. De hecho, hay descubrimientos recientes que indican que ciertas células responden solamente a palabras concretas. ¿Quizá esto vaya en el camino de la verificación de la hipótesis de la célula de la abuela?

Evidentemente hay mucho que aprender sobre el procesamiento detallado que lleva a cabo el cerebro. Se conoce muy poco, por el momento, sobre cómo llevan a cabo sus tareas los centros superiores del cerebro. Dejemos ahora esta cuestión y volvamos nuestra atención a las células reales del cerebro que le capacitan para conseguir estas notables hazañas.

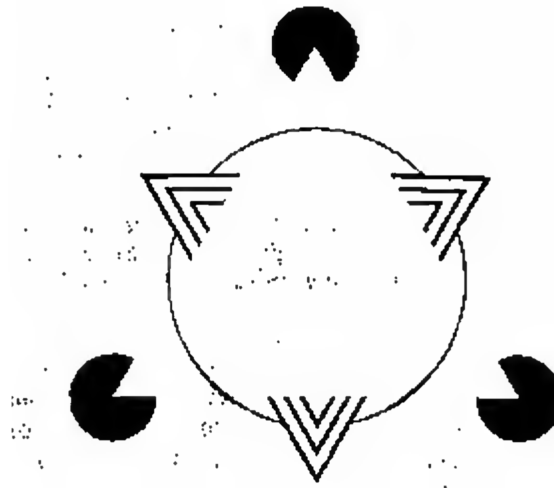


FIGURA IX.7. *¿Puede ver un triángulo blanco superpuesto a otro triángulo al cual queda sujeto mediante un anillo? El contorno del triángulo blanco no está dibujado pero aún así existen células en el cerebro que responden a estas líneas invisibles aunque percibidas.*

¿CÓMO FUNCIONAN LA SEÑALES NERVIOSAS?

Todo el procesamiento que se hace en el cerebro (y también en la médula espinal y en la retina) se logra mediante las notablemente versátiles células del cuerpo denominadas *neuronas.*[5] Tratemos de ver cómo es una neurona. En la fig. IX.8 está la imagen de una neurona. Existe un bulbo central, quizá algo parecido a una estrella, a menudo de una forma más bien similar a la de un rábano, llamado *soma,* que contiene el núcleo de la célula. De un extremo del soma parte una larga fibra nerviosa —a veces realmente muy larga (con frecuencia llega a tener varios centímetros de longitud, en un ser humano), si se tiene en cuenta que nos estamos refiriendo a

[4] *Véase* Hubel (1988) p. 221. Experimentos anteriores registraron células sensibles sólo a la imagen de una mano.
[5] La teoría, ahora bien establecida, de que el sistema nervioso consiste en células individuales separadas, las neuronas, fue vigorosamente expuesta por el gran neuroanatomista español Ramón y Cajal alrededor de 1900.

una simple célula microscópica— conocida como *axón*. El axón es el "cable" a lo largo del cual se transmite la señal *output* de la célula. El axón tiene varias bifurcaciones de las que pueden salir muchas ramas más pequeñas. Al final de cada una de estas fibras nerviosas resultantes se encuentra un pequeño *botón sinóptico*. En el otro extremo del soma, y a menudo ramificándose en todas direcciones a partir de él, existen *dendritas* arboriformes a lo largo de las cuales los datos *input* llegan al soma. (En ocasiones también hay botones sinápticos en las dendritas, dando lo que se llaman sinapsis *dendodendríticas* entre dendritas. No las tendré en cuenta en mi discusión ya que la complicación que introducen no es esencial.)

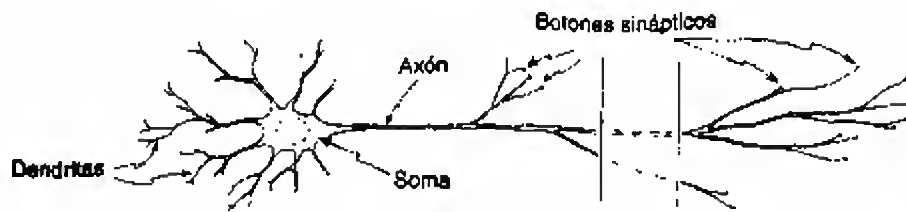


FIGURA IX.8. *Una neurona (con frecuencia es relativamente mucho más larga de lo que se indica). Hay una gran variedad de apariencias distintas para los diferentes tipos de neuronas.*

La célula entera, al ser una unidad autocontenida, tiene una membrana celular que rodea al soma, axón, botones sinápticos, dendritas y todo lo demás. Para que las señales pasen de una neurona a otra es necesario que de algún modo "salten la barrera" entre ellas. Esto se hace en una unión conocida como *sinapsis,* en donde un botón sináptico de una neurona se une a un punto de otra neurona, ya sea en el propio soma de la neurona o en una de sus dendritas (véase fig. IX.9). En realidad hay un espacio de separación muy estrecho entre el botón sináptico y el soma o dendrita a la que se une, llamado *espacio sináptico* (véase fig. IX. 10).

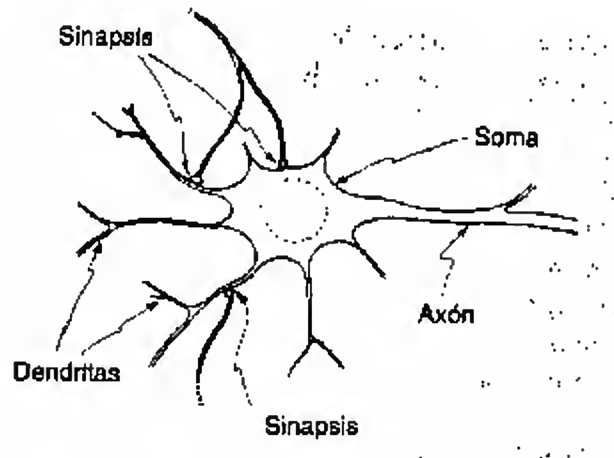


FIGURA IX.9. *Sinapsis: las uniones entre una neurona y la próxima.*

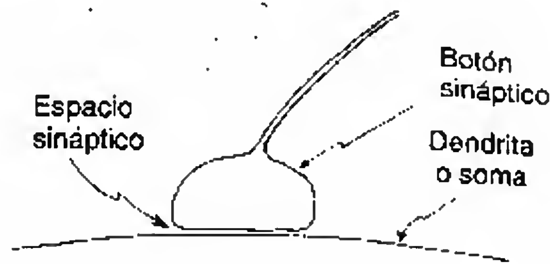


FIGURA IX.10. *Una sinapsis con más detalle. Hay una estrecha separación a través de la que fluyen los neurotransmisores químicos.*

La señal se propaga de una neurona a la siguiente a través de esta separación.

¿Qué forma adoptan las señales cuando se propagan a lo largo de las fibras nerviosas y a través de los espacios sinápticos? ¿Qué es lo que provoca que la próxima neurona emita una señal? Para un profano como yo, los procedimientos que ha adoptado de hecho la naturaleza parecen extraordinarios ¡y enormemente fascinantes! Podríamos haber pensado que las señales serían precisamente como corrientes eléctricas viajando por cables, pero es mucho más complicado que eso.

Una fibra nerviosa consiste básicamente en un tubo cilíndrico que contiene una mezcla de sal ordinaria: cloruro sódico y cloruro potásico, principalmente de este último, de modo que hay iones de sodio, potasio y cloro dentro del tubo (fig. IX. 11). Estos iones también están presentes fuera del tubo, aunque en proporciones diferentes, de modo que en el exterior hay más iones sodio que iones potasio.

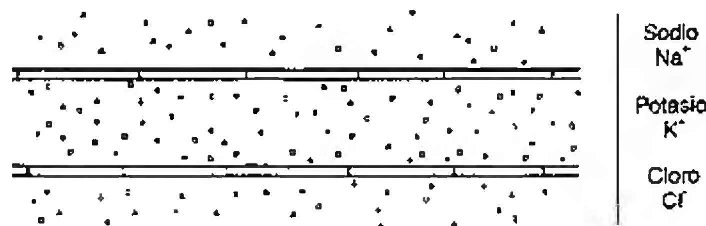


FIGURA IX.11. *Representación esquemática de una fibra nerviosa. En el estado de reposo existe en el interior un exceso de iones cloro sobre iones sodio y potasio juntos, dando lugar a una carga negativa; en el exterior sucede lo contrario, dando lugar a una carga positiva. El balance sodio/potasio es también diferente en el interior que en el exterior, con más potasio en el interior y más sodio en el exterior.*

En el estado de reposo del nervio existe una carga neta negativa en el interior del tubo (es decir, más iones cloro que iones sodio y potasio juntos, recuérdese que los iones de sodio y potasio están cargados positivamente, y los iones cloro negativamente) y una carga neta positiva en el exterior (es decir, más iones sodio y potasio que iones cloro). La membrana celular que constituye la superficie del cilindro es algo "porosa", de modo que los iones tienden a migrar a través de ella y neutralizar la diferencia de carga. Para compensar esto y mantener el exceso de carga negativa en el interior, existe una "bomba metabólica" que bombea iones sodio hacia afuera muy lentamente a través de la membrana circundante. Esto sirve también, en parte, para mantener el exceso de potasio sobre sodio en el interior. Existe otra bomba metabólica que (a

una escala algo menor) bombea iones potasio desde el exterior, y contribuye en consecuencia al exceso de potasio en el interior (aunque trabaja en contra manteniendo el desequilibrio de carga).

Una *señal* a lo largo de la fibra consiste en una región en la que este desequilibrio de carga está *invertido* (es decir, ahora es positiva en el interior y negativa en el exterior) moviéndose a lo largo de la fibra (fig. IX. 12). Imaginémonos nosotros mismos situados en la fibra nerviosa en un punto por delante de semejante región de inversión de carga. A medida que la región se aproxima, su campo eléctrico provoca que se abran pequeñas "puertas", llamadas *compuertas de sodio,* en la membrana celular; esto permite que los iones de sodio fluyan a través de ellas en dirección contraria (por una combinación de fuerzas eléctricas y presiones debidas a las diferencias de concentración, es decir, la "osmosis") desde el exterior hacia el interior. Esto da como resultado que la carga se vuelve positiva en el interior y negativa en el exterior. Cuando ha sucedido esto, la región con inversión de carga que constituye la señal nos ha alcanzado. Esto provoca ahora que se abra otro conjunto de pequeñas "puertas"*(compuertas de potasio)* que permiten que los iones potasio fluyan desde el interior y comience a restaurarse así el exceso de carga negativa en el interior. La señal ha pasado.

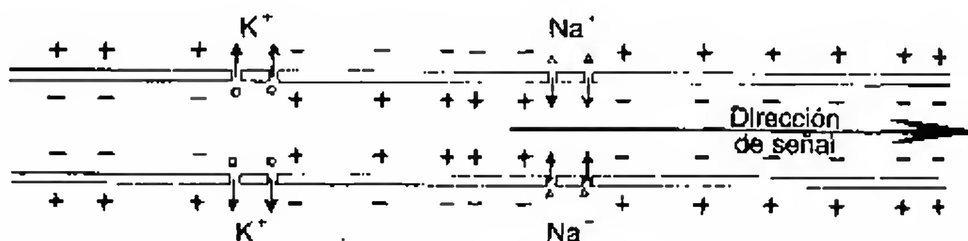


FIGURA IX. 12 *Una señal nerviosa es una región con inversión de carga que viaja a lo largo de la fibra. En su parte delantera las compuertas de sodio se abren para permitir el paso del sodio hacia el interior; y en su parte trasera, las compuertas de potasio se abren para permitir que el potasio vaya hacia afuera. Las bombas metabólicas actúan para restaurar el* statu quo.

Finalmente, a medida que la señal se aleja de nuevo, la lenta pero incesante acción de las bombas empuja otra vez hacia afuera a los iones sodio y hacia adentro a los iones potasio. Esto restaura el estado de reposo de la fibra nerviosa y la deja lista para otra señal.

Nótese que la señal consiste simplemente en una región con inversión de carga que se mueve a lo largo de la fibra. El *material* real (es decir, los iones) apenas se mueve: sólo cruza la membrana celular en una u otra dirección.

Este mecanismo "exótico" parece trabajar muy eficientemente. Es universalmente utilizado, tanto por vertebrados como por invertebrados, pero los vertebrados perfeccionaron una innovación adicional, a saber, el tener la fibra nerviosa rodeada de una funda aislante de una sustancia grasosa blanquecina llamada *mielina.* (Es esta funda de mielina la que da su color a la "sustancia blanca" del cerebro.) Este aislamiento posibilita que las señales nerviosas viajen sin pérdidas (entre "estaciones repetidoras") a una velocidad muy respetable: hasta 120 metros por segundo.

Cuando una señal llega a un botón sináptico, éste desprende una sustancia química conocida como neurotransmisor. Esta sustancia atraviesa el espacio sináptico hasta la otra neurona, ya sea en un punto de una de sus dendritas o en el propio soma. Ahora bien, algunas células tienen botones sinápticos que desprenden un neurotransmisor químico con tendencia a *excitar* el

"disparo" del soma de la siguiente neurona, es decir, a iniciar una nueva señal a lo largo de su axón. Estas sinapsis se llaman *excitatorias.* Otras tienden a *desanimar* a la siguiente neurona de que se dispare y se llaman *inhibitorias.* El efecto total de las sinapsis excitatorias que están activas en cualquier momento se suma, y de ello se resta el total de las sinapsis inhibitorias activas; si el resultado neto alcanza cierto umbral crítico, la siguiente neurona es inducida al disparo. (Las excitatorias dan lugar a una *diferencia de potencial* eléctrico positiva entre el interior y el exterior de la neurona siguiente, y las inhibitorias a una diferencia de potencial negativa. Estas diferencias de potencial se suman de la forma apropiada. La neurona se disparará cuando esta diferencia de potencial alcanza un nivel crítico en el axón asociado, de modo que el potasio no pueda salir con suficiente rapidez para restaurar el equilibrio.)

MODELOS DE COMPUTADORA

Una característica importante de la transmisión nerviosa es que las señales son (en su mayoría) fenómenos del tipo "todo o nada". La intensidad de la señal no varía: está o no está. Esto da a la acción del sistema nervioso un aspecto parecido al de una computadora. En realidad hay un gran número de semejanzas entre la acción de un gran número de neuronas interconectadas y el funcionamiento interno de una computadora digital, con sus cables conductores de corriente y sus puertas lógicas (en un momento veremos más sobre esto). No sería difícil, en principio, hacer una simulación por computadora de la acción de un sistema semejante de neuronas. Surge naturalmente una pregunta: ¿no significa esto que, cualquiera que pueda ser el cableado del cerebro, siempre podrá ser modelado por la acción de una computadora?

Para hacer más clara esta comparación diré lo que es realmente una *puerta lógica.* En una computadora tenemos también una situación "todo o nada" en donde o hay un pulso de corriente en un cable conductor o no lo hay, siendo siempre igual la intensidad del pulso cuando *está* presente. Puesto que todo esta cronometrado de manera muy exacta, la *ausencia* de un pulso será una señal definida, y será algo que pueda ser "notado" por la computadora. De hecho, cuando utilizamos el término "puerta lógica" estamos considerando implícitamente que la presencia o ausencia de un pulso significa "verdadero" o "falso" respectivamente. En realidad, esto no tiene nada que ver con la verdad o falsedad real: sólo se utiliza normalmente para hacer comprensible una terminología. Escribamos también el dígito "1" para *"verdadero"* (presencia de pulso) y "0" para *"falso"* (ausencia de pulso) y utilicemos "&" para "y" (que es el "enunciado" de que ambos son "verdaderos", es decir, la respuesta es 1 si y sólo si ambos argumentos son 1), "V" para "o" (que "significa" que uno u otro o ambos son "verdaderos", es decir, 0 si y sólo si ambos argumentos son 0), "=> " para "implica" (es decir, A => B significa que "si A es verdadero entonces B es verdadero", que es equivalente a "o A es falso o B es verdadero"), "<=>" para "si y solo si" (ambos "verdaderos" o ambos "falsos"), y "~" para "no" ("verdadero" si "falso"; "falso" si "verdadero"). Podemos describir la acción de las diversas operaciones lógicas en términos de las llamadas "tablas de verdad":

$$A\&B: \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad AVB: \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\text{A=>B:}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} \qquad \text{A<=>B:}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

donde, en cada caso, A etiqueta las filas (esto es, A = 0 da la primera fila y A = 1, la segunda) y análogamente B etiqueta las columnas. Por ejemplo, si A = 0 y B = 1, lo que da la entrada superior derecha de cada tabla, entonces en la *tercera* tabla tenemos 1 para el valor de A => B. (Un ejemplo verbal de esto en términos de lógica *real:* la afirmación "si estoy dormido entonces soy feliz" se verifica ciertamente —de manera trivial— en el caso particular en que resulte estar a la vez despierto y feliz.). Finalmente, la puerta lógica "no", sencillamente tiene el efecto:

$$\sim 0=1 \quad \text{y} \quad \sim 1=0.$$

Estos son los tipos básicos de puertas lógicas. Existen algunas otras, pero todas éstas pueden construirse a partir de las mencionadas.[6]

Ahora bien, ¿podemos en principio construir una computadora a base de conexiones de *neuronas?* Indicaré que, incluso sólo con las consideraciones muy primitivas sobre el disparo de las neuronas que acabamos de discutir, esto es realmente posible. Veamos cómo puede ser posible en principio construir puertas lógicas a base de conexiones de neuronas. Necesitamos una nueva manera de codificar los dígitos, puesto que la *ausencia* de una señal no desencadena nada. Consideremos (de forma arbitraria) que un pulso *doble* denota 1 (o "verdadero") y un pulso *simple* denota O (o "falso") y consideremos un esquema simplificado en el que el umbral de disparo de una neurona sea siempre igual a *dos* pulsos excitatorios simultáneos. Es fácil construir una puerta "y" (es decir, "&"). Como se muestra en la fig. IX. 13, podemos considerar que las dos fibras nerviosas de entrada terminan en el único par de botones sinápticos en la neurona de salida. (Si ambos son pulsos dobles, entonces tanto el primer pulso como el segundo alcanzarán el umbral exigido de dos pulsos, mientras que si uno de ellos es un pulso simple, entonces sólo uno del par alcanzará este umbral. Estoy suponiendo que los pulsos están minuciosamente cronometrados y que, para más concreción, en el caso de un pulso doble es el *primero* del par el que determina el cronometraje.) La construcción de una puerta "no" (es decir, "~") es mucho más complicada, y una manera de hacerlo se da en la fig. IX. 14. Aquí la señal de entrada procede a lo largo de un axón que se divide en dos ramas. Una rama toma una ruta circular, de longitud tal como para retardar la señal en un intervalo exactamente igual al intervalo de tiempo entre los dos pulsos de un pulso doble; luego ambas ramas se bifurcan una vez más, con una rama de cada bifurcación terminando en una neurona inhibitoria, pero de forma que la que viene de la rama retrasada se desdobla primero para dar una ruta directa y otra circular.

[6] De hecho, *todas* las puertas lógicas pueden construirse a partir de "~" y "&", o incluso sólo a partir de *una sola* operación ~(A & B).

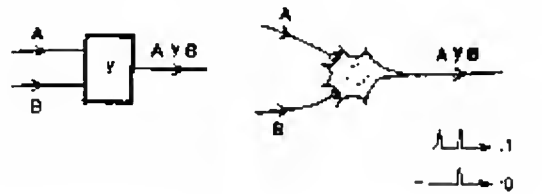


FIGURA IX.13. *Una puerta "y"- En el "modelo de neurona" de la derecha, se considera que la neurona se "dispara" sólo cuando el* input *alcanza una intensidad doble de la de un pulso simple.*

La salida de dicha neurona sería *nada,* en el caso de un pulso simple de entrada, y *un pulso doble* (en posición retardada) en el caso de un pulso doble de entrada. El axón que conduce esta salida se desdobla en tres ramas, todas ellas terminando en botones sinápticos inhibitorios de una neurona final excitatoria. Cada una de las dos ramas restantes del axón originalmente dividido se divide una vez más en dos y las cuatro ramas terminan también en esta neurona final pero ahora con botones sinápticos excitatorios. El lector puede verificar que esta neurona excitatoria final proporciona la salida requerida "no" (es decir, un pulso doble si la entrada es uno simple y un pulso simple si la entrada es uno doble). (Este esquema parece absurdamente complicado, pero es lo mejor que he podido imaginar.)

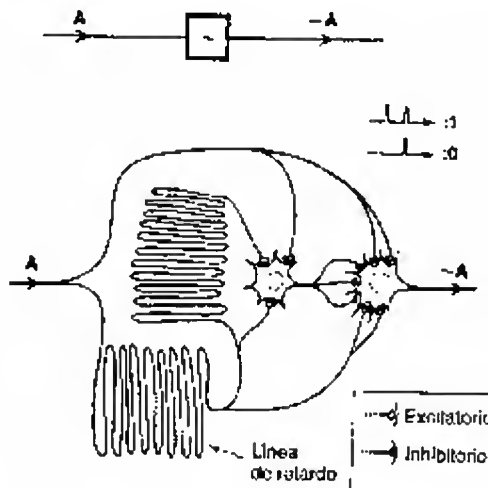


FIGURA IX. 14. *Una puerta "no". En el "modelo de neurona" se necesita de nuevo un* input *de una intensidad doble (al menos) para que una neurona se dispare.*

El lector o lectora puede entretenerse buscando construcciones "neuronales" directas para las otras puertas lógicas citadas.

Por supuesto, estos ejemplos explícitos no deben ser considerados como modelos serios de lo que es el cerebro en detalle. Sólo estoy tratando de indicar que existe una equivalencia lógica esencial entre el modelo de disparo de neurona que he dado arriba y la construcción de una computadora electrónica. Es fácil ver que una computadora puede simular cualquier modelo semejante de interconexiones neuronales; a su vez, las construcciones detalladas dadas arriba dan una indicación del hecho de que, recíprocamente, los sistemas de neuronas pueden simular una computadora y, por lo tanto *podrían* actuar como una máquina (universal) de Turing. Aunque la

discusión de las máquinas de Turing que se dio en el capítulo II no utilizaba puertas lógicas,[7] y de hecho necesitamos mucho más que sólo puertas lógicas si tenemos que simular una máquina de Turing general, no existen nuevas cuestiones de principio implicados en su realización —*siempre que* nos esté permitido *aproximar* la *cinta infinita* de una máquina de Turing por un banco de neuronas grande pero finito. Esto parecería ser un argumento de que los cerebros y las computadoras son esencialmente equivalentes.

Antes de dar un salto demasiado apresurado hacia esta conclusión, consideraremos diversas diferencias entre la acción cerebral y la acción de las computadoras actuales. En primer lugar, he simplificado bastante mi descripción del disparo de una neurona como un fenómeno todo-o-nada. Eso se refiere a un pulso simple que viaja a lo largo de un axón pero, en realidad, cuando una neurona se "dispara" emite toda una secuencia de tales pulsos en rápida sucesión. Incluso cuando una neurona no está activada, emite pulsos pero sólo a un ritmo lento. Cuando se dispara es la *frecuencia* de estos pulsos sucesivos la que aumenta enormemente. Existe también un aspecto probabilístico en el disparo de una neurona. El mismo estímulo no siempre produce el mismo resultado. Además, la acción cerebral no tiene exactamente el cronometraje preciso que se necesita para las corrientes de una computadora electrónica; y habría que señalar que la acción de las neuronas —a un ritmo máximo de unas 1000 veces por segundo— es mucho más lenta que la de los circuitos electrónicos más rápidos, en un factor de aproximadamente 10^{-6}.

También, a diferencia del cableado muy preciso de una computadora electrónica, parece que hubiera mucha aleatoriedad y redundancia en la forma detallada en que están conectadas las neuronas; aunque, a diferencia de lo que se creía hace cincuenta años, hoy sabemos que la precisión en el cableado del cerebro es considerable desde el nacimiento.

Casi todo lo anterior parecería ser una desventaja para el cerebro comparado con una computadora. Pero hay otros factores a favor del cerebro. Con las puertas lógicas hay muy pocos cables de entrada y de salida (digamos tres o cuatro como mucho) mientras que las neuronas pueden tener un número enorme de sinapsis. (Como un ejemplo extremo, las neuronas del cerebelo conocidas como células de Purkinje tienen unos 80000 terminales sinápticas excitatorias.) Además, el número total de neuronas del cerebro todavía supera con mucho al número de transistores incluso en la mayor computadora actual: probablemente 10^{11} para el cerebro y "sólo" unas 10^{9} para la computadora. Pero, por supuesto, es muy probable que la cifra para la computadora aumente en el futuro.[8] Además, el alto número de células cerebrales aparece generalmente debido al inmenso número de pequeñas *células granulosas* que se encuentran en el cerebelo: alrededor de treinta mil millones (3×10^{10}). Si creemos que es simplemente el enorme número de neuronas el que nos permite tener experiencias conscientes, mientras que las actuales computadoras no parecen tenerlas, entonces necesitamos encontrar alguna explicación adicional de por qué la acción del *cerebelo* parece ser completamente inconsciente, mientras que la conciencia puede estar asociada con el *cerebro* propiamente dicho, que tiene sólo el doble de neuronas (unas 7×10^{10}) con una densidad mucho más pequeña.

---

7 En realidad, el uso de puertas lógicas está más próximo a la construcción de una computadora electrónica de lo que lo están las consideraciones detalladas de la máquina de Turing del capítulo II. En este capítulo se hacía hincapié en la aproximación de Turing por razones teóricas. El desarrollo de las computadoras reales surgió tanto del trabajo del matemático húngaro-estadounidense John von Neumann como del de Alan Turing.

8 Estas comparaciones son equívocas en ciertos aspectos. La inmensa mayoría de los transistores en las computadoras actuales concierne a la "memoria" más que a la acción lógica, y la memoria de la computadora siempre puede ser aumentada por medios exteriores, indefinidamente en la práctica. Con la operación paralela aumentada, más transistores estarían involucrados directamente en la acción lógica de lo que sucede normalmente en la actualidad.

PLASTICIDAD CEREBRAL

Hay algunos otros puntos de diferencia entre la acción cerebral y la acción de una computadora que me parecen ser de mucha más importancia que los mencionados hasta ahora, y tienen que ver con un fenómeno conocido como *plasticidad cerebral.* No es realmente legítimo considerar el cerebro simplemente como una colección *fija* de cables neuronales. Las interconexiones entre neuronas no son fijas, como lo serían en el modelo de computadora anterior, sino que están cambiando continuamente. No quiero decir que cambien las localizaciones de los axones o las dendritas; gran parte de su complejo "cableado" está establecido en líneas generales desde el nacimiento. Me estoy refiriendo a las uniones sinápticas en las que tiene lugar realmente la comunicación entre neuronas diferentes. A menudo estas ocurren en lugares llamados *espinas dendríticas,* que son minúsculas protuberancias en las dendritas con las cuales pueden tomar contacto los botones sinápticos (véase fig IX 15). Aquí, "contacto" ya *no* significa simplemente tocar sino dejar una separación (espacio sináptico) de la distancia justa, alrededor de cuarenta milésimas de milímetro. En ciertas condiciones, estas espinas dendríticas pueden contraerse y romper el contacto, o pueden (ellas u otras nuevas) crecer y hacer nuevos contactos. Por lo tanto, si pensamos que las conexiones de neuronas en el cerebro constituyen en efecto una computadora, entonces es una computadora capaz de cambiar continuamente.

Según una de las principales teorías sobre cómo se establece la memoria a largo plazo, tales cambios en las conexiones sinápticas son los que proporcionan los medios para almacenar la información necesaria. Si esto es así, vemos entonces que la plasticidad cerebral no es ya una complicación accidental sino que es una característica *esencial* de la actividad del cerebro.

¿Cuál es el mecanismo subyacente a estos cambios continuos? ¿Con qué rapidez se efectúan estos cambios? La respuesta a la segunda pregunta parece controvertida, pero hay al menos una escuela de pensamiento que mantiene que tales cambios sólo pueden ocurrir en unos pocos segundos. Eso habría que esperar si tales cambios son los responsables del almacenamiento de memorias permanentes, puesto que tales memorias pueden asentarse en cuestión de segundos *(cfr.* Kandel, 1976). Esto tendrá significativas consecuencias para nosotros más adelante. Volveré a esta importante cuestión en el próximo capítulo.

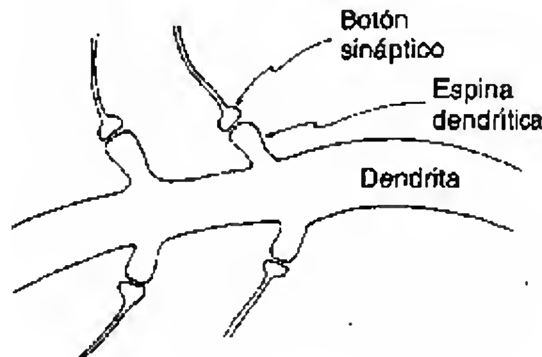


FIGURA. IX 15. *Uniones sinápticas que incluyen espinas dendríticas. La efectividad de la unión queda inmediatamente afectada por el crecimiento o contracción de la espina.*

¿Qué hay de los mecanismos subyacentes a la plasticidad cerebral? Una ingeniosa teoría (debida a Donald Hebb, 1954) propone la existencia de ciertas sinapsis (ahora llamadas "sinapsis de Hebb") con la propiedad siguiente: una sinapsis de Hebb entre una neurona A y una neurona B será reforzada cuando el disparo de A vaya seguido del disparo de B, y debilitada si no lo hace. Esto es independiente de si la propia sinapsis de Hebb está o no involucrada de modo significativo en causar el disparo de B. Ello da lugar a cierta forma de "aprendizaje". Se han propuesto varios modelos matemáticos para tratar de simular una actividad de aprendizaje de solución de problemas basada en este tipo de teoría. Estos se conocen como *redes murales.* Parece que tales modelos son realmente capaces de algún tipo de aprendizaje rudimentario, pero hasta ahora están lejos de ser modelos realistas del cerebro. En cualquier caso, parece probable que los mecanismos que controlan los cambios en las conexiones sinápticas sean bastante más complicados que los considerados.

En relación con esto existe otro aspecto de la liberación de neurotransmisores por los botones sinápticos. A veces ésta no tiene lugar en los espacios sinápticos sino que los neurotransmisores entran en el fluido intercelular general, quizá para influir a otras neuronas a gran distancia. Muchas sustancias neuroquímicas diferentes parecen emitirse de esta forma, y existen teorías de la memoria diferentes de la que he citado, que dependen de las posibles variedades distintas de tales sustancias químicas que puedan estar involucradas. Ciertamente el estado del cerebro puede ser influido de un modo general por la presencia de sustancias químicas que son producidas por otras partes del cerebro *(v.g.* como sucede con las hormonas). El problema general de la neuroquímica es complicado, y es difícil ver cómo proporcionar una simulación por computadora detallada y fiable de todo lo que pueda ser importante.

COMPUTADORAS PARALELAS Y LA "UNICIDAD" DE LA CONCIENCIA

Mucha gente parece ser de la opinión de que el desarrollo de las computadoras *paralelas* tiene la clave para construir una máquina con las capacidades de un cerebro humano. Consideraremos a continuación esta idea muy popular actualmente. Una computadora en paralelo, como opuesta a una computadora en serie, lleva a cabo simultáneamente un gran número de cálculos independientes, y los resultados de estas operaciones básicamente autónomas son combinados intermitentemente para dar contribuciones al cálculo global. La motivación para este tipo de computadora proviene básicamente de un intento de imitar el modo de operar del sistema nervioso, puesto que diferentes partes del cerebro parecen llevar a cabo funciones de cálculo separadas e independientes *(v g* el procesamiento de información visual en la corteza visual).

Hay que hacer aquí dos puntualizaciones. La primera es que no hay diferencia *en principio* entre una computadora en serie y una en paralelo. Ambas son de hecho *máquinas de Turing (cfr.* capítulo II. Sólo puede haber diferencias en eficiencia, o rapidez, del cálculo como un todo. Hay algunos tipos de cálculo para los que una organización en paralelo es realmente más eficiente, pero no es éste el caso siempre ni mucho menos. El segundo punto es que, al menos en mi opinión, es muy poco probable que la computación en paralelo clásica tenga la clave de cómo se está desarrollando nuestro pensamiento *consciente.* Una característica distintiva del pensamiento consciente (al menos cuando estamos en un estado psicológico normal, y no sometidos a una

operación de "escisión cerebral") es su "unicidad", en oposición a las muchas actividades independientes que transcurren al mismo tiempo.

Expresiones tales como "¿cómo puede esperar usted que piense más de una cosa cada vez?" son un lugar común. ¿Es *siquiera* posible mantener cosas separadas en la conciencia simultáneamente? Quizá se *puedan* mantener unas pocas cosas a la vez, pero esto se parece más a un continuo ir y venir entre los varios temas sobre los que se está pensando simultánea, consciente e independientemente. Si fuéramos a pensar conscientemente sobre dos cosas totalmente independientes sería más probable que tuviésemos dos *conciencias separadas,* incluso aunque fuera sólo por un corto tiempo, mientras que parece que experimentamos (al menos en una persona normal) *una sola* conciencia que puede ser vagamente consciente de cierto número de cosas, pero que en un instante dado se concentra en *una sola* cosa concreta.

Por supuesto, lo que queremos decir aquí con "una sola cosa" no está completamente claro. En el próximo capítulo encontraremos algunos ejemplos muy notables de "pensamientos simples" en las inspiraciones de Poincaré y Mozart. Pero no tenemos que ir tan lejos para reconocer que aquello de lo que una persona puede ser consciente en un instante dado puede ser implícitamente muy complicado. Imaginemos que vamos a decidir lo que queremos cenar, por ejemplo. Podrá haber mucha información implícita en una idea consciente semejante, y una descripción verbal completa sería bastante larga.

Esta "unicidad" de la percepción consciente me parece que está en fuerte oposición con la imagen de una computadora paralela. Por el contrario, esta imagen podría ser más apropiada como un modelo de la acción *inconsciente* del cerebro. Se pueden realizar simultáneamente y de forma más *o* menos automática varios movimientos independientes diferentes —andar, abrochar un botón, respirar, o incluso hablar— sin tener una conciencia absoluta de *ninguno* de ellos.

Por el contrario, es posible que pudiera existir alguna relación entre esta "unicidad" de la conciencia y el *paralelismo cuántico.* Recuérdese que, según la teoría cuántica, está permitido que coexistan en una superposición opciones diferentes en el nivel cuántico. Por lo tanto, un *estado cuántico simple* podría consistir en principio en un gran número de actividades diferentes, todas ellas ocurriendo simultáneamente. Esto es lo que quiere decir el paralelismo cuántico, y en un momento consideraremos la idea teórica de una "computadora cuántica" en la que podría utilizarse en principio este paralelismo para ejecutar un gran número de cálculos simultáneos. Si un "estado mental" consciente pudiera ser en algún modo semejante a un estado cuántico, entonces alguna forma de "unicidad" o globalidad de pensamiento parecería más apropiada de lo que lo sería en el caso de una computadora paralela ordinaria. Hay algunos aspectos atractivos en esta idea a los que volveré en el próximo capítulo. Pero antes de que podamos mantener seriamente esta idea debemos plantear la cuestión de si es probable que los efectos cuánticos tengan importancia en la actividad del cerebro.

¿HAY UN PAPEL PARA LA MECÁNICA CUÁNTICA EN LA ACTIVIDAD CEREBRAL

Las discusiones anteriores sobre la actividad cerebral han sido totalmente clásicas, excepto en el momento en que ha sido necesario apelar a fenómenos físicos cuyas causas implícitas subyacentes son parcialmente mecánico-cuánticas *(v.g.* iones, con su carga eléctrica, compuertas de sodio y de potasio, los potenciales químicos definidos que determinan el carácter sí/no de las

señales nerviosas, la química de los neurotransmisores). ¿Hay algún papel más definido para un control genuinamente mecánico-cuántico en algún lugar clave? Esto parecería necesario si la discusión del final del capítulo anterior fuera a tener auténtica importancia.

Existe, de hecho, al menos un lugar evidente en donde la acción en el nivel mecánico-cuántico simple puede tener importancia para la actividad neuronal, y éste está en la *retina.* (Recuérdese que técnicamente la retina es parte del cerebro.) Experimentos con sapos han mostrado que, en condiciones adecuadas, *un solo fotón* que incida en la retina adaptada a la oscuridad puede ser suficiente para desencadenar una señal nerviosa macroscópica (Baylor, Lamb y Yau, 1979) Lo mismo parece ser cierto para el hombre (Hecht, Shlaer y Pirenne, 1941) pero en este caso hay presente un mecanismo adicional que suprime tales señales débiles de modo que no hagan confusa la imagen percibida con demasiado "ruido" visual. Se necesita una señal combinada de unos *siete* fotones para que un humano adaptado a la oscuridad puede hacerse consciente realmente de su llegada. De todas formas, parece que en la retina humana hay células con sensibilidad a un solo fotón.

Puesto que en el cuerpo humano *existen* neuronas que pueden ser disparadas por sucesos cuánticos simples, ¿no es razonable preguntar si podrían encontrarse células de este tipo en algún lugar de las partes principales del cerebro humano? Por lo que yo conozco no existe evidencia de esto. Todos los tipos de célula que se han examinado requieren que se alcance un umbral, y se necesita un número muy elevado de cuantos para que se dispare la célula. Podríamos especular, no obstante, que en algún lugar profundo del cerebro se encontrarán células con sensibilidad a un solo cuanto. Si se probara que este es el caso, entonces la mecánica cuántica estaría involucrada de manera significativa en la actividad cerebral.

Incluso esto no parece *útil,* puesto que el cuanto se está utilizando simplemente como medio de desencadenar una señal y no se ha obtenido ningún efecto característico de interferencia cuántica. Hasta ahora, todo lo que logremos será una incertidumbre sobre si una neurona se disparará o no, y esto no será de mucha ayuda para nosotros.

Sin embargo, volvamos a considerar la retina. Supongamos que llega un fotón a la retina, habiendo sido reflejado previamente en un espejo semirreflectante. Su estado implica una superposición lineal compleja de su impacto en una célula de la retina y su no impacto en una célula de la retina, saliendo en lugar de ello, por ejemplo, por la ventana hacia el espacio *(cfr.* fig. VI. 17). Cuando llega el instante en el *que pudiera* haber golpeado la retina, y en tanto que la regla lineal U de la teoría cuántica siga siendo válida (es decir, evolución determinista del vector de estado de acuerdo con la ecuación de Schrödinger), tendremos una superposición lineal compleja de una señal nerviosa y una ausencia de señal nerviosa. Cuando ésta incida sobre la conciencia del sujeto sólo se percibirá *una* de estas opciones teniendo lugar efectivamente, y se deberá haber realizado el otro procedimiento mecánico-cuántico **R** (reducción del vector de estado). (Al decir esto estoy dejando de lado la idea de los muchos universos, que tiene su multitud de problemas propios.) En línea con las consideraciones apuntadas al final del capítulo anterior deberíamos preguntar si el paso de la señal perturba la cantidad de materia suficiente para que se alcance el criterio de *un gravitón.* Aunque es cierto que en la retina se consigue una impresionante amplificación para convertir la energía del fotón en un movimiento de masa en la señal real —quizá en un factor que llegue a 10^{20} en masa movida— esta masa queda aún muy lejos de la masa de Planck m_p en un factor muy grande (digamos del orden de 10^8). Sin embargo, una señal nerviosa crea un *campo eléctrico* variable detectable en sus proximidades (un campo

toroidal, con el nervio corno eje, moviéndose a lo largo del nervio). Este campo podría perturbar significativamente el *entorno* y el criterio de un gravitón podría alcanzarse fácilmente en dicho entorno. Por lo tanto, según el punto de vista que he estado exponiendo, el procedimiento **R** podría haberse efectuado ya antes de que percibamos o no el destello de luz, según sea el caso. Desde este punto de vista nuestra conciencia no es necesaria para reducir el vector de estado.

COMPUTADORAS CUÁNTICAS

Si *especulamos* con que dichas neuronas sensibles a un solo cuanto están jugando un papel importante en lo más profundo del cerebro, podemos preguntar qué efectos tendrán. Discutiré primero el concepto de Deutsch de *computadora cuántica (cfr.* también capítulo IV) y luego plantearé si podemos considerarlo importante para nuestras reflexiones sobre este punto.

Como se indicó antes, la idea básica consiste en hacer uso del paralelismo cuántico según el cual debe considerarse que dos cosas muy diferentes tienen lugar simultáneamente en una superposición lineal cuántica, como el fotón que simultáneamente se refleja y atraviesa el espejo semirreflectante, o quizá atraviesa cada una de las dos rendijas. Para una computadora cuántica estas dos diferentes cosas superpuestas serían, en su lugar, dos *computaciones* diferentes. No nos suponemos interesados en obtener las respuestas a *ambos* cómputos, sino en algo que utilice parte de la información extraída del par superpuesto. Finalmente, se haría una "observación" adecuada sobre el par de cómputos, cuando ambos estuvieran acabados, para obtener la respuesta requerida.[9] De esta forma, el dispositivo podría ahorrar tiempo ejecutando los dos cómputos simultáneamente. Hasta aquí no parece que haya una ganancia significativa al proceder de esta manera, ya que presumiblemente sería mucho más directo utilizar un par de computadoras clásicas distintas en paralelo (o una sola computadora paralela clásica) que una computadora cuántica. Sin embargo, la ganancia real para una computadora cuántica llegaría cuando se necesitase un número *muy grande* de cálculos paralelos —quizá un número indefinidamente grande— cuyas respuestas no nos interesasen individualmente, sino sólo alguna combinación apropiada de todos los resultados.

En concreto, la construcción detallada de una computadora cuántica implicaría una versión cuántica de una puerta lógica, en donde el *output* sería el resultado de cierta "operación unitaria" aplicada al *input* —un caso de la acción de **U** — y toda la operación de la computadora estaría llevando a cabo hasta el final un proceso **U**, hasta que un "acto de observación" final haría entrar en acción a **R.**

Según el análisis de Deutsch, las computadoras cuánticas no pueden utilizarse para ejecutar operaciones no-algorítmicas (es decir, cosas más allá del poder de una máquina de Turing), pero pueden, en ciertas situaciones muy artificiales, obtener una velocidad mayor, en el sentido de la *teoría de la complejidad*, que la de una máquina de Turing estándar. Hasta aquí estos resultados son un poco decepcionantes para una idea tan sorprendente, pero todavía estamos en el principio.

¿Qué relación podría tener esto con la acción de un cerebro que contenga un número significativo de neuronas sensibles a un solo cuanto? El mayor problema con la analogía sería

---

[9] Deutsch prefiere utilizar, en sus descripciones, el punto de vista de los "muchos universos" con respecto a la teoría cuántica. Sin embargo, es importante darse cuenta de que esto es completamente accesorio, y que el concepto de computadora cuántica es igualmente apropiado cualquiera que sea el punto de vista que adoptemos con respecto a la mecánica cuántica estándar.

que los efectos cuánticos se perderían rápidamente en el "ruido", el cerebro es un objeto demasiado "caliente" para preservar la coherencia cuántica (es decir, el comportamiento usualmente descrito por la acción continuada de **U**) durante cualquier intervalo de tiempo apreciable. En mis propios términos, esto significaría que continuamente se estaría cumpliendo el criterio de un-gravitón, de modo que la operación R continuaría todo el tiempo entremezclada con **U**.

Todo esto no parece muy prometedor si esperamos sacar de la mecánica cuántica algo útil para el cerebro. Quizá estemos condenados a ser computadoras después de todo. Yo, personalmente, no lo creo. Pero se necesitan consideraciones adicionales si queremos descubrir nuestra vía de escape.

¿MAS ALLA DE LA TEORIA CUÁNTICA?

Quiero volver a un punto que ha sido un tema subyacente en gran parte de este libro. ¿Nuestra imagen de un mundo gobernado por las reglas de las teorías clásica y cuántica, tal como las entendemos actualmente, es verdaderamente adecuada para la descripción de cerebros y mentes.

Ciertamente existe un interrogante para cualquier descripción cuántica "ordinaria" de nuestros cerebros, ya que la acción de "observación" se considera un ingrediente esencial de la interpretación válida de la teoría cuántica convencional. ¿Debe considerarse el cerebro "observándose a sí mismo" cuando un pensamiento o percepción emerge a la atención consciente? La teoría convencional nos deja sin ninguna regla clara de cómo la mecánica cuántica podría tomar esto en consideración y, en consecuencia, aplicarlo al cerebro como un todo. He intentado formular un criterio para la aparición de R que es independiente de la conciencia (el "criterio de un gravitón") y si se pudiera desarrollar algo parecido a esto en una teoría completa coherente entonces podría emerger una manera de proporcionar una descripción cuántica del cerebro más clara de la que existe actualmente.

Sin embargo, creo que estos problemas fundamentales no surgen sólo en nuestros intentos de describir la acción cerebral. Las acciones de los propias computadoras digitales dependen de forma vital de efectos cuánticos, que, en mi opinión, no están totalmente libres de las dificultades inherentes a la teoría cuántica. ¿Cuál es esta dependencia cuántica "vital"? Para comprender el papel de la mecánica cuántica en el cálculo digital debemos preguntar primero cómo podríamos intentar hacer que un objeto totalmente *clásico* se comporte como una computadora digital. En el capítulo V considerábamos la "computadora de bolas de billar" clásica de Fredkin-Toffoli; pero señalábamos también que este "dispositivo" teórico depende de ciertas idealizaciones que dejan de lado un problema de inestabilidad esencial inherente a los sistemas clásicos. Este problema de inestabilidad se describió como una difusión efectiva en el espacio de fases a medida que evoluciona el tiempo (fig. V.14), lo que conduce a una casi inevitable pérdida de precisión en la operación de un dispositivo clásico. Es la mecánica cuántica la que en última instancia detiene esta degradación de la precisión. En las computadoras electrónicas modernas es necesaria la existencia de *estados discretos* (por ejemplo, para codificar los dígitos 0 y 1), de modo que la cuestión de si la computadora se encuentra en uno de estos estados o en el otro tiene una respuesta precisa. Esto constituye la propia esencia de la naturaleza "digital" de la operación de la computadora. Esta discrecionalidad depende en última instancia de la mecánica cuántica. (Recordemos la discrecionalidad cuántica de los estados de energía, de las frecuencias

espectrales, del *spin,* etc., *cfr.* capítulo VI.) Incluso las viejas calculadoras mecánicas dependían de la *solidez* de sus diversas partes, y la solidez, también, descansa realmente en la discrecionalidad de la teoría cuántica.[10]

Pero la discrecionalidad cuántica no es obtenible de la sola acción de **U.** La ecuación de Schrödinger es *peor* aún para impedir la indeseable difusión y "pérdida de precisión" de lo que lo son las ecuaciones de la física clásica. Según **U**, una función de onda de una simple partícula, inicialmente localizada en el espacio, se difundirá ella misma sobre regiones cada vez mayores a medida que pasa el tiempo. Los sistemas más complejos también estarían sometidos a veces a esta irrazonable falta de localización (recuérdese el gato de Schrödinger) si no fuera por la acción ocasional de **R.** (Los estados *discretos* de un átomo, por ejemplo, son aquellos con energía, momento y momento angular total definidos. Un estado general que "se difunde" es una superposición de estados discretos semejantes. Es la acción de R, en alguna etapa, la que impone que el átomo "esté" realmente en uno de estos estados discretos.)

Ni la mecánica clásica ni la cuántica —esta última sin ciertos cambios fundamentales que colocaran a R en un proceso "real"— podrán explicar la forma en que *pensamos.* Quizá incluso la acción digital de las computadoras necesita una comprensión más profunda de la interrelación entre las acciones de **U** y **R**. Con las computadoras sabemos, al menos, que la acción es *algorítmica* (porque así los hemos diseñado), y no tratamos de sacar provecho de ningún supuesto comportamiento *no* algorítmico en las leyes físicas. Pero yo afirmo que con los cerebros y las mentes la situación es muy diferente. Se puede alegar plausiblemente que existe un ingrediente *no* algorítmico esencial en los procesos de pensamiento (consciente). En el próximo capítulo expondré las razones de mi creencia en tal ingrediente y especularé acerca de los efectos físicos reales que constituir una "conciencia" que influya en la acción del cerebro.

[10] Este comentario podría no ser aplicable si se consideran los constituyentes "clásicos" como ejes, levas, etc. Pienso en tales constituyentes como partículas ordinarias (digamos esféricas o puntuales).

X. ¿DÓNDE RESIDE LA FÍSICA DE LA MENTE?

¿PARA QUÉ SON LAS MENTES?

CUANDO SE DISCUTE EL PROBLEMA MENTE-CUERPO se suele centrar la atención en dos temas: "¿cómo es posible que una conciencia pueda *surgir* realmente a partir de un objeto material (un cerebro)?"; y recíprocamente "¿cómo es posible que una conciencia, mediante la acción de la voluntad, *influya* realmente en el movimiento (que en apariencia está determinado físicamente) de los objetos materiales?" Estos son los aspectos pasivo y activo del problema mente-cuerpo. Parece que tenemos, en la "mente" (o más bien en la "conciencia"), "algo" no material que, por una parte, está producido por el mundo material y, por otra, puede influirlo. Sin embargo, en mis discusiones preliminares en este último capítulo preferiré considerar un problema algo diferente y quizá más científico —que tiene importancia tanto para el problema pasivo como para el activo— con la esperanza de que nuestros intentos de encontrar una respuesta puedan acercarnos un poco hacia una mejor comprensión de estos viejos enigmas fundamentales de la filosofía. Mi pregunta es: ¿qué *ventaja selectiva* confiere la conciencia a quienes realmente la poseen?

Al enunciar la pregunta de esta manera se hacen implícitamente varias suposiciones. En primer lugar está la creencia de que la conciencia es realmente "algo" *descriptible científicamente.* Se supone también que esta "cosa" realmente "hace algo" y, además, que lo que hace es útil para la criatura que la posee, de modo que otra cualquiera criatura equivalente, aunque sin conciencia, se comportaría de una forma menos eficaz. En el extremo opuesto se podría pensar que la conciencia es simplemente un concomitante pasivo de la posesión de un sistema de control suficientemente elaborado, y que *no* "hace" *nada* realmente por sí sola. (Este último sería presumiblemente el punto de vista de los defensores de la IA fuerte, por ejemplo.) Como alternativa, quizá exista algún propósito divino o misterioso para el fenómeno de la conciencia —posiblemente un propósito teleológico aún no revelado a nosotros— y cualquier discusión de este fenómeno simplemente en términos de selección natural pasaría por alto este "propósito". Algo preferible, para mi modo de pensar, sería una versión más científica de este tipo de argumento, a saber, el *principio antrópico,* que afirma que la naturaleza del Universo en el que nosotros mismos nos encontramos está fuertemente condicionada por la exigencia de que deben estar presentes seres sensibles como nosotros para observarla. (Se aludió brevemente a este principio en el capítulo VIII, y volveré a él más tarde.)

Abordaré la mayoría de estos temas a su debido tiempo, pero notemos primero que el término "mente" es quizá un poco equívoco cuando nos referimos al problema mente-cuerpo. Después de todo, hablamos a menudo de "la mente inconsciente". Esto demuestra que no consideramos que los términos "mente" y "conciencia" sean sinónimos. Quizá cuando nos referimos a la mente inconsciente tenemos una vaga imagen de "alguien ahí detrás" que actúa entre bastidores pero que normalmente (excepto quizá en sueños, alucinaciones, obsesiones o lapsus freudianos) no tiene incidencia directa sobre lo que percibimos. Quizá la mente inconsciente *tenga* realmente una conciencia autónoma, pero esta conciencia se mantiene normalmente separada de la parte de la mente a la que normalmente nos referimos como "nosotros".

Puede que esto no sea tan rebuscado como parece a primera vista. Existen experimentos que parecen indicar que puede haber algún tipo de conciencia presente aun cuando un paciente esté siendo operado con anestesia general; en el sentido de que conversaciones mantenidas en ese momento pueden influir "inconscientemente" más tarde sobre el paciente, y pueden ser después recordadas bajo hipnosis como si realmente hubieran sido experimentadas en su momento.

Además, algunas sensaciones que parecían haber quedado bloqueadas para la conciencia mediante sugestión hipnótica pueden ser recordadas más tarde bajo nueva hipnosis como si "hubieran sido experimentadas", aunque en una "vía diferente" *(cfr.* Oakley y Eames, 1985). Estos resultados no están nada claros para mí, aunque no concibo que fuera correcto asignar un conocimiento ordinario a la mente inconsciente, y no tengo deseos de tratar aquí estas especulaciones. De todas formas, la división entre la mente consciente y la inconsciente es ciertamente un tema sutil y complicado al que tendremos que volver.

Tratemos de ser tan claros como sea posible sobre lo que entendemos por conciencia y cuándo creemos que está presente. No creo que sea prudente, en este estadio de comprensión, intentar proponer una *definición* precisa de conciencia, pero podemos fiarnos en buena medida de lo que nuestras impresiones subjetivas y sentido común intuitivo nos dicen sobre el significado del término y sobre cuándo es probable que esté presente esta propiedad de conciencia. Yo sé, más o menos, cuándo estoy consciente, y supongo que los demás experimentan algo análogo. Para estar consciente, parece que debo ser consciente *de* algo, quizá de una sensación como dolor o calor, o de una escena animada, o de un sonido musical; o quizá soy consciente de un sentimiento tal como intriga, desesperación o felicidad; o puedo ser consciente del recuerdo de alguna experiencia pasada, o de llegar a una comprensión de lo que algún otro está diciendo, o de una nueva idea propia; o puedo estar intentando conscientemente hablar o ejecutar alguna otra acción como la de levantarme de mi asiento. También puedo "remontarme al pasado" y ser consciente *de* tales intenciones o de mi sentimiento de dolor, o de mi experiencia de un recuerdo o de llegar a una comprensión; o incluso puedo ser simplemente consciente de mi propia conciencia. Puedo estar dormido y ser aún consciente en cierto grado, siempre que tenga algún sueño; o quizá, a medida que me voy despertando, estoy influyendo conscientemente en el rumbo de ese sueño. Estoy dispuesto a creer que la conciencia es una cuestión de grado y no simplemente algo que está o no está. Supondré que la palabra "conciencia" es esencialmente sinónimo de "conocimiento" (aunque quizá conocimiento es algo más pasivo que lo que entiendo por conciencia), mientras que "mente" y "alma" tienen connotaciones adicionales que son mucho menos definibles en el presente. Tendremos suficientes dificultades para entender lo que es la conciencia tal como está, así que espero que el lector me perdonará si no me ocupo en esencia de los problemas adicionales de "mente" y "alma".

Está también la cuestión de qué entendemos por el término "inteligencia". Esto es, después de todo, lo que interesa a la gente de la IA, antes que el tema quizá más difuso de la "conciencia". Alan Turing (1950), en su famoso artículo *(cfr.* capítulo I), no se refería directamente a la "conciencia" sino al "pensamiento", y en el título figuraba la palabra "inteligencia". En mi propia forma de ver las cosas, la cuestión de la inteligencia es subsidiaria de la de la conciencia. Me parece poco concebible que la verdadera inteligencia pudiera estar presente a menos que estuviera acompañada de la conciencia. Por el contrario, si resultara que la gente de la IA *fueran* finalmente capaces de simular inteligencia sin que la conciencia esté presente, entonces podría considerarse insatisfactorio definir el término "inteligencia" sin incluir esta inteligencia simulada. En ese caso, el tema de la "inteligencia" no me interesaría realmente para este punto. Estoy interesado principalmente en la conciencia.

Cuando afirmo mi propia creencia en que la verdadera inteligencia requiere conciencia, estoy sugiriendo implícitamente (puesto que yo no creo en la tesis de la IA fuerte de que la simple activación de un algoritmo produciría la conciencia) que la inteligencia no puede simularse adecuadamente mediante procedimientos algorítmicos, es decir, mediante una computadora, en

el sentido en que hoy utilizamos el término. (Véase la discusión de la "Prueba de Turing" en el capítulo I.) En efecto, argumentaré con fuerza dentro de un momento (véase concretamente la discusión del pensamiento matemático dada tres secciones mas adelante) que debe haber un ingrediente esencialmente no algorítmico en la actuación de la conciencia.

Abordemos a continuación la cuestión de si *existe* una diferencia operacional entre algo que sea consciente y alguna cosa "equivalente" no lo sea. ¿Revelaría siempre su presencia la conciencia en algún objeto? Me gustaría pensar que la respuesta a esta pregunta es necesariamente "sí". Sin embargo, mi fe en esto se ve alentada por la falta total de consenso sobre en qué partes del reino animal debe encontrarse la conciencia. Algunos no admiten que pueda poseerla en absoluto algún animal no humano (y, algunos, ni siquiera que la poseyeran los seres humanos antes de alrededor del año 1000 a. C., *cfr.* Jaynes, 1980), mientras que otros atribuirán conciencia a un insecto, un gusano, o incluso a una roca! Por mi parte dudaría que un gusano o un insecto —aunque ciertamente no una roca— tenga mucho, si tiene algo, de esta cualidad; pero los mamíferos, de modo general, me dan la impresión de tener cierta genuina conciencia. De esta falta de consenso podemos inferir, al menos, que no hay un criterio generalmente aceptado para la manifestación de la conciencia. Pudiera ser todavía que *exista* una impronta de comportamiento consciente, aunque no universalmente reconocida. Aun así, esto sólo podría significar el papel *activo* de la conciencia. Es difícil ver cómo la simple presencia de conciencia, sin su contrapartida activa, pudiera ser directamente verificada. Esto tiene un apoyo en el terrible hecho de que, durante algún tiempo en los años cuarenta, el curare fue utilizado como "anestésico" en operaciones realizadas a niños, cuando en realidad el efecto de esta droga es paralizar la acción de los nervios motores sobre los músculos, de modo que la agonía que realmente experimentaban estos infortunados niños no tenía modo de percibirla el cirujano en esa época *(cfr.* Dennett, 1978).

Volvamos al posible papel activo que pueda tener la conciencia. ¿Es necesario que la conciencia deba jugar —y a veces lo hace realmente— un papel activo discernible operacionalmente? Tengo diversas razones para creerlo. Una primera razón la constituye el modo por el que, utilizando el "sentido común", tenemos a menudo la sensación de que percibimos directamente que alguna otra persona *está* realmente consciente? No parece muy probable que *esa* impresión sea falsa.* Mientras que puede darse el caso de que una persona que *está* consciente no lo esté de modo obvio (como los niños drogados con curare), no es probable que una persona que no está consciente parezca estarlo. Por lo tanto debe haber realmente algún modo de comportamiento que es característico de la conciencia (incluso aunque no sea *siempre* manifestado por la conciencia) y al que somos sensibles por medio de nuestras "intuiciones de sentido común".

En segundo lugar, consideremos el implacable proceso de la selección natural. Veamos este proceso a la luz del hecho de que, como hemos visto en el anterior capítulo, no toda la actividad del cerebro es directamente accesible a la conciencia. De hecho, el cerebelo "más antiguo" —con su enorme superioridad de densidad local de neuronas— parece llevar a cabo acciones muy complejas sin que la conciencia esté en modo alguno involucrada directamente. No obstante, la naturaleza ha decidido que evolucionen seres sensibles como nosotros, en lugar de contentarse con criaturas que podrían estar dirigidas por mecanismos de control completamente inconscientes. Si la conciencia no sirve para ningún propósito selectivo, ¿por qué la naturaleza se

* Al menos con la tecnología de computadora actual (véase la discusión respecto a la prueba de Turing en el capítulo I).

tomó la molestia de hacer evolucionar cerebros conscientes cuando parece que hubieran bastado cerebros "autómatas" no sensibles como los cerebelos?

Además, existe una sencilla "razón definitiva" para creer que la conciencia debe tener algún efecto activo, incluso si este efecto *no* tiene ninguna ventaja selectiva. En efecto, ¿por qué los seres como nosotros nos preocupamos a veces —especialmente cuando se investiga en el tema— por cuestiones sobre el "yo"? (Casi podría decir: "¿por qué están *ustedes* leyendo este capítulo?" o "¿por qué sentí *yo* en primer lugar un fuerte deseo de escribir un libro sobre este tema?") Es difícil imaginar que un autómata completamente inconsciente perdiera su tiempo en estos temas. Puesto que, por el contrario, los seres conscientes *parecen* actuar de esta extraña manera de vez en cuando, ellos se están comportando en consecuencia de una forma que es *diferente* de la forma en que lo harían si *no fueran* conscientes, así que la conciencia tiene *algún* efecto activo. Por supuesto, no habría problema en programar deliberadamente una computadora para que pareciera que se comportaba de esta forma ridícula *(v.g.* podría ser programada para que deambulara murmurando: "Oh, ¡Dios mío¡, ¿cuál es el sentido de la vida?" "¿Por qué estoy aquí?" "¿Qué diantres es este 'yo' que siento?"). Pero ¿por qué se molestaría la selección natural en favorecer semejante raza de individuos, cuando seguramente el implacable mercado libre de la jungla debería haber extirpado hace tiempo tales absurdos inútiles?

Me parece claro que estas meditaciones y murmuraciones a que nos entregamos cuando (quizá temporalmente) nos hacemos filósofos no son cosas que sean seleccionadas *por sí mismas,* sino que son el "equipaje" necesario (desde el punto de vista de la selección natural) que deben llevar los seres que *son* conscientes y cuya conciencia ha sido elegida por la selección natural, aunque por alguna razón diferente y presumiblemente muy poderosa. Es un equipaje que no es demasiado gravoso y es fácil de llevar (aunque quizá a regañadientes), yo diría, para las indómitas fuerzas de la selección natural. En ocasiones, tal vez cuando se dan la paz y prosperidad de que a veces disfruta nuestra afortunada especie, y no tenemos que estar siempre luchando contra los elementos (o nuestros vecinos) por la supervivencia, los tesoros del contenido de nuestro equipaje pueden empezar a intrigarnos y asombrarnos. Es cuando vemos que otros se comportan con esta extraña conducta filosófica cuando nos quedamos *convencidos* de que estamos tratando con individuos, distintos de uno mismo, que también tienen mentes.

¿QUÉ HACE REALMENTE LA CONCIENCIA?

Aceptemos que la presencia de conciencia en una criatura supone realmente una ventaja selectiva para la misma. ¿Cuál sería en concreto esta ventaja? Una idea que he oído expresar es que la conciencia podría ser ventajosa para un cazador que trata de imaginar lo que probablemente haría su presa a continuación, "poniéndose él mismo en lugar" de la presa. Imaginarse que él mismo *es* la presa podría darle una ventaja sobre ella.

Es perfectamente posible que haya parte de verdad en esta idea, pero no me siento muy a gusto con ella. En primer lugar, ello supone alguna conciencia preexistente por parte de la misma presa, puesto que difícilmente sería útil imaginar que uno "sea" un autómata, ya que un autómata —inconsciente por definición— no es algo que *se pueda* "ser" conscientemente. En cualquier caso, también podríamos imaginar que un cazador autómata totalmente inconsciente pudiera contener en sí mismo, como parte de su programa, una subrutina que fuera el programa real de su

presa autómata. No me parece que sea *lógicamente* necesario que la conciencia tenga que entrar de ninguna forma en esta interrelación cazador-presa.

Por supuesto, es difícil ver cómo el proceso aleatorio de la selección natural podría haber sido suficientemente ingenioso para dar a un cazador *autómata* una copia completa del programa de la presa. Esto sonaría más a *espionaje* que a selección natural. Y un programa *parcial* (en el sentido de un trozo de "cinta" de una máquina de Turing, o algo que se le aproxime) apenas sería ventaja selectiva para el cazador. Parecería necesaria la improbable posesión de toda la cinta, o al menos de toda una parte autocontenida en ella. En consecuencia, y como una posible alternativa, podría haber una verdad parcial en la idea de que algún elemento de conciencia, y no simplemente un programa de computadora, pueda inferirse del argumento cazador-presa. Pero ello no parece abordar el *verdadero* problema de cuál es realmente la diferencia entre una acción consciente y una "programada".

La idea aludida arriba parece estar relacionada con un punto de vista sobre la conciencia que se oye exponer a menudo, a saber: que un sistema tendrá "conciencia" de alguna cosa si tiene dentro de sí un modelo de esa cosa, y que se hace *auto*consciente cuando tiene dentro de sí un modelo de *sí mismo.* Pero un programa de computadora que contenga dentro de sí (digamos como subrutina) alguna descripción de otro programa no hace al primero consciente del segundo; ni ningún aspecto *auto*referencial le hace *auto*consciente. A pesar de las afirmaciones que parecen hacerse con frecuencia, los temas reales concernientes a la conciencia y autoconciencia apenas se tocan, en mi opinión, en consideraciones de este tipo. Una videocámara no es consciente de las escenas que está registrando; y tampoco una videocámara que esté dirigida hacia un espejo posee autoconciencia. (fig. X.1).

Deseo seguir una línea diferente. Hemos visto que no todas las actividades que llevan a cabo nuestros cerebros están acompañadas de atención consciente (y, en particular, la acción cerebelar no parece que sea consciente). ¿Qué podemos hacer con el pensamiento consciente que no podamos hacer inconscientemente? El problema se hace más escurridizo por el hecho de que cualquier cosa que hagamos parece exigir conciencia inicialmente, pues parece también capaz de ser aprendida y luego ejecutada inconscientemente (quizá por el cerebelo). De algún modo se necesita la conciencia para manejar situaciones en las que tenemos que hacer nuevos juicios, y en las que no se han establecido reglas por

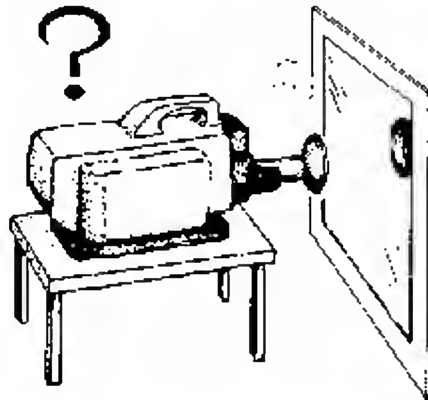


FIGURA X. 1. *Una videocámara dirigida hacia un espejo se forma un modelo de ella dentro de sí misma. ¿Le hace esto autoconsciente?*

adelantado. Es difícil ser muy preciso sobre las distinciones entre los tipos de actividad mental que parecen exigir la conciencia y las que no lo hacen. Quizá, como sostendrían los defensores de la IA fuerte (y otros), nuestra "formación de nuevos juicios" estaría aplicando de nuevo algunas reglas algorítmicas bien definidas, pero unas de un oscuro "alto nivel" de cuyo funcionamiento no somos conscientes. Sin embargo, creo que el tipo de terminología que tendemos a utilizar, que distingue nuestra actividad mental consciente de la inconsciente, al menos *sugiere* una distinción no algorítmica:

Conciencia necesaria	Conciencia no necesaria
"sentido común"	"automático"
"juicio de verdad"	"seguir reglas despreocupadamente"
"comprensión"	"programado"
"valoración artística"	"algorítmico"

Quizá estas distinciones no sean siempre muy precisas, debido en particular a que intervienen muchos factores inconscientes en nuestros juicios conscientes: experiencia, intuición, prejuicio, incluso nuestra utilización normal de la lógica. Pero, afirmaré, los propios juicios son las manifestaciones de la actuación de la *conciencia.* Sugiero así que, mientras que las acciones inconscientes del cerebro son las que proceden según procesos algorítmicos, la acción de la conciencia es muy diferente y actúa de una forma que no puede describirse mediante ningún algoritmo.

Resulta irónico que las ideas que estoy exponiendo aquí representan casi una inversión de otras que he oído frecuentemente. A menudo se dice que es la mente *consciente* la que se comporta de la forma "racional" que podemos entender, mientras que el inconsciente es el misterioso. La gente que trabaja en la IA afirma a menudo que en cuanto podamos comprender conscientemente alguna línea de pensamiento también podremos ver cómo hacer una computadora capaz de seguirla; es de los misteriosos procesos inconscientes de los que no tenemos (¡aún!) idea de cómo tratar. Mi propia línea argumental ha sido que los procesos inconscientes podría perfectamente ser algorítmicos, pero en un nivel muy complicado que sería monstruosamente difícil desentrañar en detalle. El pensamiento completamente consciente que puede racionalizarse como algo enteramente lógico puede otra vez (a menudo) formalizarse como algo algorítmico, pero esto se hace a un *nivel completamente diferente.* No estamos pensando ahora en las operaciones internas (disparo de neuronas, etc.) sino en la manipulación de pensamientos globales. A veces, esta manipulación mental tiene un carácter algorítmico (como en la lógica primitiva: los antiguos silogismos griegos tal y como fueron formalizados por Aristóteles o la lógica simbólica del matemático George Boole; *cfr.* Gardner, 1958); y a veces no lo tiene (como en el teorema de Gödel y algunos de los ejemplos dados en el capítulo IV). La *formación de juicios,* que afirmo es la impronta de la conciencia, es *ella misma* algo sobre lo que la gente de la IA no tendría ninguna idea de cómo programar en una computadora.

Se objeta a veces que, después de todo, los *criterios* para estos juicios no son conscientes, de modo que ¿por qué estoy atribuyendo estos juicios a la conciencia? Esto sería no captar el punto esencial de las ideas que estoy tratando de expresar. No estoy pidiendo que comprendamos conscientemente *cómo* formamos nuestras impresiones y juicios conscientes; eso sería caer en la confusión de niveles a la que me acabo de referir. Las *razones* subyacentes a nuestras

impresiones conscientes serían cosas no directamente accesibles a la conciencia. Éstas tendrían que considerarse en un nivel físico más profundo que el de los pensamientos reales de que tenemos conciencia. (Haré un intento en una hipótesis posterior.) Son las propias impresiones conscientes las que *son* los juicios (no algorítmicos).

De hecho, en los últimos capítulos ha estado subyacente la cuestión de que parece haber algo *no algorítmico* en nuestro pensamiento consciente. En particular, una consecuencia del argumento del capítulo IV concerniente en especial al teorema de Gödel, era que, al menos en matemáticas, la contemplación consciente puede a veces capacitarnos para comprobar la verdad de un enunciado de un modo en que no podría hacerlo un algoritmo. (Elaboraré este argumento un poco más adelante.) En realidad, los algoritmos por sí mismos *nunca* comprueban la verdad. Sería tan fácil hacer que un algoritmo sólo produjera falsedades como hacer que produjera verdades. Necesitamos *intuiciones externas* para decidir la validez o no de un algoritmo (más adelante volveremos sobre esto). Estoy exponiendo aquí el argumento de que es esta capacidad para distinguir (o "intuir"), en circunstancias apropiadas, verdad de falsedad (y belleza de fealdad) lo que constituye la impronta de la conciencia.

Aclararé, sin embargo, que no estoy dando a entender ninguna forma de mágica "adivinanza". La conciencia no es de ninguna ayuda al tratar de conjeturar el número de la suerte de una lotería (limpia). Me estoy refiriendo a los juicios que hacemos continuamente mientras estamos conscientes, reuniendo todos los hechos, impresiones sensoriales y experiencias recordadas que son de importancia, y sopesando las cosas, incluso formando en ocasiones juicios inspirados. En principio hay información disponible suficiente para que se pueda hacer el juicio importante, pero el proceso de formular el juicio apropiado, extrayendo lo que se necesita del amasijo de datos, puede ser algo para lo que no existe ningún procedimiento algorítmico evidente o incluso, en caso de que exista uno, puede que no sea práctico. Quizá tengamos una situación que, una vez hecho el juicio, puede haber algo más que un proceso algorítmico (o quizá simplemente uno más fácil) en *verificar* que el juicio sea exacto en la formación inicial de dicho juicio. Estoy conjeturando que en tales circunstancias, la conciencia seguiría su propósito mediante una forma de conjurar los juicios apropiados.

¿Por qué digo que la impronta de la conciencia es una formación de juicios no algorítmica? La razón procede en parte de mi experiencia como matemático. Sencillamente no confío en mis acciones algorítmicas inconscientes cuando mi conciencia no les ha prestado la atención debida. Con frecuencia no hay nada falso en el algoritmo *como tal* cuando se ha realizado algún cálculo, pero ¿era el algoritmo *correcto* a elegir para el problema a resolver? En un sencillo ejemplo tendríamos que aprender las reglas algorítmicas para multiplicar dos números y también para dividir un número por otro (o podemos preferir recurrir a la ayuda de una calculadora de bolsillo algorítmica), pero ¿cómo sabemos si, para el problema a resolver, tendríamos que multiplicar o dividir los números? Para ello necesitamos *pensar* y hacer un juicio *consciente.* (Veremos dentro de un momento por qué tales juicios deben ser, al menos algunas veces, no algorítmicos.) Por supuesto, una vez que hemos tratado un gran número de problemas similares, la decisión sobre si multiplicar o dividir los números puede convertirse en una segunda naturaleza y puede ser ejecutada de manera algorítmica, quizá por el cerebelo. En esta etapa ya no es necesaria la conciencia, y nos llegamos a sentir confiados para permitir a nuestra mente consciente vagar y contemplar otras materias — aunque, de cuando en cuando, podemos tener necesidad de verificar que el algoritmo no se ha desviado de alguna forma (quizá sutil).

Cosas de este mismo tipo están sucediendo continuamente en todos los niveles del pensamiento matemático. A menudo, cuando hacemos matemáticas, nos esforzamos por conseguir nuevos algoritmos, pero el propio intento no parece ser un procedimiento algorítmico. Una vez que se encuentra un algoritmo apropiado el problema está resuelto en cierto sentido. Además, el juicio matemático sobre si algún algoritmo es realmente preciso o apropiado es el tipo de cosas que requiere mucha atención consciente. Algo similar ocurría en la discusión de los sistemas formales para las matemáticas que se describieron en el capítulo IV. Podemos empezar con algunos axiomas de los que deben derivarse diversas proposiciones matemáticas. El último procedimiento puede ser realmente algorítmico, pero es necesario que se haga un juicio por un matemático consciente para decidir si los axiomas son apropiados. El que estos juicios son necesariamente *no* algorítmicos debería hacerse más evidente a partir de la discusión que se hace dos secciones más adelante. Pero antes de llegar a esto consideraremos lo que podría ser un punto de vista predominante sobre lo que nuestros cerebros están haciendo y cómo ha sido posible.

¿SELECCIÓN NATURAL DE ALGORITMOS?

Si suponemos que la acción del cerebro humano, consciente o no, consiste simplemente en la ejecución de algún algoritmo muy complicado, entonces debemos preguntar cómo se formó un algoritmo de eficacia tan extraordinaria. La respuesta normal, por supuesto, sería que surgió por "selección natural". Por lo que respecta a las criaturas con cerebros evolucionados, aquellos con los algoritmos más eficaces tendrían más probabilidades de sobrevivir y por ello, en general, tendrían más progenie. Esta progenie también tendería a portar algoritmos más eficaces que sus primos, puesto que heredaron de sus padres los ingredientes de estos mejores algoritmos; de este modo los algoritmos mejoraron poco a poco —no necesariamente de manera uniforme puesto que pudo haber tropiezos importantes en su evolución— hasta que alcanzaron el importante estado que (aparentemente) encontramos en el cerebro humano. *(cfr.* Dawkins, 1986.)

Incluso según mi propio punto de vista tendría que haber algo de verdad en esa imagen, puesto que concibo que gran parte de la acción del cerebro es realmente algorítmica y, como el lector habrá deducido de lo anterior, creo firmemente en el poder de la selección natural. Pero no veo cómo la selección natural por sí sola pueda hacer evolucionar algoritmos que pudieran tener el tipo de juicios conscientes sobre la *validez* de otros algoritmos que al parecer tenemos.

Imaginemos un programa ordinario de computadora. ¿Cómo llegó a formarse? Es evidente que no (directamente) por selección natural. Algún programador humano de computadoras lo habrá concebido, verificando que realiza correctamente las acciones que se supone debe hacer. (En realidad, muchos programas de computadora complicados contienen errores —normalmente menores, pero a menudo muy sutiles y que no salen a la luz excepto en circunstancias muy poco comunes. La presencia de tales errores no afecta medularmente a mi argumento.) A veces un programa de computadora puede haber sido "escrito" por otro programa, digamos un programa de computadora "maestro", pero en tal caso el propio programa maestro habrá sido el producto del ingenio y la intuición humanos; o el programa podría perfectamente ensamblarse a partir de ingredientes, algunos de los cuales son los productos de otros programas de computadora. Pero en todos los casos la validez y la misma concepción del programa habrá sido en última instancia responsabilidad de (al menos) una conciencia humana.

Podemos imaginar, por supuesto, que no es necesario que haya sido así y que, dado el tiempo suficiente, el programa de computadora pudo haber evolucionado espontáneamente por algún proceso de selección natural. Si creemos que las acciones de las conciencias de los programadores de computadoras son en sí mismas simples algoritmos, entonces debemos creer que los algoritmos han evolucionado de esta misma forma. Lo que me molesta de esto, sin embargo, es que la decisión sobre la validez de un algoritmo *no* es en sí misma un proceso algorítmico. Ya hemos visto algo de esto en el capítulo II. (La cuestión de si una máquina de Turing se *parará* o no, es un punto que no puede decidirse algorítmicamente.) Para decidir si un algoritmo *funcionará* o no, necesitamos *perspicacia,* y no sólo otro algoritmo.

De todas formas, aun sería posible imaginar algún tipo de proceso de selección natural que fuera efectivo para producir algoritmos *aproximadamente* válidos. Sin embargo, yo personalmente encuentro esto muy difícil de creer. Cualquier proceso de selección natural de este tipo actuaría sólo sobre el *output* de los algoritmos* y no directamente sobre las ideas inherentes a los algoritmos. Esto no sólo es extremadamente ineficiente; creo que sería totalmente impracticable. En primer lugar, no es fácil verificar cuál es realmente un algoritmo mediante un simple examen de su *output.* (Sería bastante sencillo construir dos acciones simples y muy diferentes de máquina de Turing para las que las cintas de salida no difieran hasta, digamos, el lugar 2^{65536}, diferencia que no se podría reconocer en toda la historia del Universo.) Además, la más ligera "mutación" de un algoritmo —por ejemplo, un pequeño cambio en la especificación de una máquina de Turing o en su cinta de *input*— podría hacerla totalmente inútil, y es difícil ver siquiera cómo de esta forma aleatoria podrían aparecer *mejoras* reales en los algoritmos. (Incluso las mejoras *deliberadas* son difíciles sin que estén disponibles los "significados". Esto se confirma por los casos no poco frecuentes en los que un Programa de computadora complicado y mal documentado necesita ser alterado o corregido y el programador original se ha marchado o quizá ha muerto. Antes que tratar de desentrañar todos los diversos significados e intenciones de los que el programa depende explícitamente, probablemente sea más fácil desecharlo sin más y empezar todo de nuevo.)

Quizá pudiera imaginarse una forma mucho más "robusta" de especificar algoritmos que no estuviera sujeta a las críticas anteriores. En cierto modo, esto es lo que yo afirmo. Las especificaciones "robustas" son las *ideas* inherentes a los algoritmos. Pero las ideas son cosas que, por lo que sabemos, necesitan mentes conscientes para manifestarse. Hemos vuelto al problema de qué es la conciencia y qué cosas puede hacer de las que son incapaces los objetos inconscientes, y cómo la selección natural ha podido ser lo bastante inteligente para hacer evolucionar las cualidades más importantes. Los productos de la selección natural son realmente sorprendentes.

El escaso conocimiento que he adquirido acerca de cómo funcionan los cerebros humanos —y, de hecho, cualquier otra cosa viviente— me deja perplejo de asombro y admiración. El funcionamiento de una neurona es extraordinario, pero las propias neuronas están organizadas en conjunto, en el instante del nacimiento, de una forma muy notable con un gran número de conexiones dispuestas para todas las tareas que sean necesarias más adelante. No es sólo la propia conciencia lo que es notable, sino todo lo que parece ser necesario para sostenerla.

*. Esta también el asunto algo espinoso de si dos algoritmos deben ser considerados equivalentes sólo con que sus *outputs* sean iguales, en lugar de serlo sus cómputos reales. Véase capítulo II.

Si alguna vez descubrimos en detalle qué permite a un objeto físico llegar a ser consciente, entonces sería concebible que pudiéramos construir tales objetos por nosotros mismos, aunque podrían no calificarse como "máquinas" en el sentido que ahora entendemos. Podríamos imaginar que estos objetos tendrían una tremenda ventaja sobre nosotros, puesto que podrían diseñarse *específicamente* para una tarea, a saber, llegar a *tener conciencia.* No tendrían que crecer a partir de una sola célula. No tendrían que arrastrar el "equipaje" de su ascendencia (las viejas e "inútiles" partes del cerebro o del cuerpo que sobreviven en nosotros gracias a los "accidentes" de nuestros remotos ancestros). Podríamos imaginar, a la vista de estas ventajas, que tales objetos podrían tener éxito en superar efectivamente a los seres humanos en las tareas en las que (en opinión de gente como yo) las computadoras algorítmicas están condenadas a la subordinación.

Pero puede ser que la conciencia no se reduzca a esto. Quizá nuestra conciencia depende de nuestra herencia y los miles de millones de años de evolución *real* que hay tras nosotros. A mi modo de ver, hay todavía algo misterioso en la evolución con su aparente "andar a tientas" hacia algún propósito. Al menos, parece que las cosas se organizan algo mejor de lo que "deberían" hacerlo sólo sobre la base de la evolución por ciego azar y selección natural. Podría perfectamente suceder que tales apariencias fueran bastante engañosas. Parece que hubiera algo en el modo de actuar de las leyes de la física que permitiera que la selección sea un proceso mucho más eficaz de lo que sería con leyes arbitrarias. El aparente "andar a tientas inteligentemente" que resulta es un tema interesante sobre el que volveré brevemente más adelante.

LA NATURALEZA NO ALGORÍTMICA DE LA PERSPICACIA MATEMÁTICA

Como he declarado antes, una buena parte de las razones para creer que la conciencia es capaz de influir en los juicios de verdad de una manera no algorítmica se derivan del teorema de Gödel. Si podemos ver que el papel de la conciencia es no algorítmico cuando formamos juicios *matemáticos,* en lo que el cálculo y la demostración rigurosa constituyen un factor tan importante, entonces podremos persuadirnos de que un ingrediente no algorítmico semejante podría ser también crucial para el papel de la conciencia en circunstancias más generales (no matemáticas).

Recordemos el argumento dado en el capítulo IV que establecería el teorema de Gödel y su relación con la computabilidad. Se demostró allí que *cualquiera* que sea el algoritmo (suficientemente extenso) que un matemático pudiera utilizar para establecer la verdad matemática —o, lo que es equivalente,[1] cualquiera que sea el *sistema formal* que pudiera adoptar para proporcionar su criterio de verdad— habría siempre proposiciones matemáticas, como las proposiciones de Gödel $P_k(k)$ explícitas del sistema para las que su algoritmo no puede proporcionar una respuesta. Si el funcionamiento de la mente matemática es completamente algorítmico, entonces el algoritmo (o sistema formal) que utilice para formar sus juicios no es capaz de tratar la proposición $P_k(k)$ construida a partir de su algoritmo personal. De todas formas podemos ver (en principio) que $P_k(k)$ es *verdadera.* Esto parecería presentarle una contradicción

[1] Ya vimos en el capítulo IV, que la comprobación de la validez de una prueba en un sistema formal es siempre algorítmica. Inversamente, todo algoritmo que genere verdades matemáticas siempre puede ser añadido a los axiomas y a las reglas de procedimientos de la lógica común ("cálculo de predicados") y conseguir así un nuevo sistema formal de obtención de verdades matemáticas.

puesto que esa mente debería ser capaz de verlo. Quizá esto indique que la mente matemática *no* estaba utilizando en absoluto un algoritmo.

Éste es esencialmente el argumento de Lucas (1961), de que la actividad del encéfalo no puede ser algorítmica enteramente, sin embargo un buen número de argumentos en contra han aparecido a través del tiempo (Benacerraf, 1967; Good, 1969; Lewis, 1969, 1989; Hofstadter, 1981, Bowie, 1982). Con relación a esta discusión, quiero destacar que los términos "algorítmico" y "no algorítmico" se refieren a cualquier cosa que (efectivamente) pueda ser simulada en una computadora de uso general. Esto incluye ciertamente "acción paralela", pero también "redes neurales" (o "máquina de conexión"), "heurística", "aprendizaje" (siempre y cuando anticipadamente se haya fijado un procedimiento sobre cómo se supone que el artefacto va a aprender) e interacción con el medio ambiente (que puede ser simulado con la cinta *input* de la máquina de Turing). El más serio de estos argumentos en contra es éste, para convencernos de la verdad de $P_k(k)$ necesitaríamos *saber* cuál es el algoritmo del matemático, y también estar convencidos de su validez como medio de llegar a la verdad matemática. Si el matemático utilizara un algoritmo muy complicado en su cabeza no tendríamos *ninguna posibilidad* de conocer realmente cuál es dicho algoritmo y, por consiguiente, no seríamos capaces de construir su proposición de Gödel, y mucho menos convencernos de su validez. Este tipo de objeción se plantea a menudo contra las afirmaciones como las que hago aquí, de que el teorema de Gödel indica que los juicios matemáticos humanos son no algorítmicos. Pero en lo personal no encuentro convincente esta objeción. Supongamos, por el momento, que los matemáticos forman sus juicios conscientes sobre la verdad matemática de una manera realmente algorítmica. Trataremos de reducir esto al absurdo *(reductio ad absurdum)* utilizando el teorema de Gödel. Debemos considerar en primer lugar la proposición de que diferentes matemáticos utilicen algoritmos *no equivalentes* para encontrar la verdad. Sin embargo, una de las características más sorprendentes de la matemática (quizá casi única entre las disciplinas) es que la verdad de las proposiciones puede realmente establecerse mediante argumentos abstractos. Un argumento matemático que convence a un matemático —siempre que no contenga error alguno— también convencerá a otro en cuanto el argumento haya sido completamente captado. Esto también se aplica a las proposiciones tipo Gödel. Si el primer matemático está dispuesto a aceptar todos los axiomas y reglas de inferencia de un sistema formal que produzca sólo proposiciones *verdaderas,* entonces también debe estar dispuesto a aceptar que su proposición de Gödel describe una proposición verdadera. Sucedería exactamente lo mismo con el segundo matemático. El punto esencial es que los argumento que establecen la verdad matemática son *comunicables.*[2]

Por lo tanto, no nos referimos a los diversos algoritmos oscuros que podrían estar rondando la mente de distintos matemáticos. Aludimos a un sistema formal empleado universalmente que es *equivalente* a todos los diferentes algoritmos de los matemáticos para juzgar la verdad

---

[2] A algunos lectores podrá inquietar el hecho de que haya diferentes puntos de vista entre los matemáticos. Recuérdese lo expuesto en el capítulo IV. Sin embargo las diferencias, donde existen, no necesariamente tienen que interesar aquí: se refieren sólo a las cuestiones esotéricas relativas a conjuntos muy grandes, así que basta con que restrinjamos nuestra atención a proposiciones de la aritmética (con un número finito de cuantificadores existenciales y universales) aplicando lo que ya se ha dicho. (Quizá esto exagera algo la situación, puesto que puede utilizarse a veces un punto de reflexión relativo a conjuntos infinitos para derivar proposiciones de la aritmética.) En cuanto al formalista muy dogmático, inmune a Gödel, que afirma no reconocer siquiera que *exista* tal cosa como una verdad matemática, simplemente lo paso por alto puesto que evidentemente no posee la virtud de adivinar la verdad que acabe la discusión.
Por supuesto, los matemáticos a veces se equivocan. Parece que el mismo Turing creía que era *aquí* donde estaba la salida a los argumentos tipo Gödel contra el pensamiento humano como desarrollo algorítmico. Pero me parece improbable que la falibilidad humana sea la clave para la reflexión. (Y lo *azaroso* puede ser simulado muy bien por medios algorítmicos.)

matemática. Ahora bien, ni siquiera puede saberse si este presunto sistema "universal", o algoritmo, es el que utilizamos los matemáticos para establecer la verdad. En efecto, si lo fuera, *podríamos* construir su proposición de Gödel y saber que es también una verdad matemática. Por lo tanto, llegamos a la conclusión de que el algoritmo que utilizan los matemáticos para establecer la verdad matemática es tan complicado u oscuro que nunca podremos conocer su propia validez.

Pero esto hace caso omiso de la esencia de la matemática. Toda la gracia de nuestra herencia y aprendizaje matemático es que *no* nos rendimos a la autoridad de reglas oscuras que nunca podemos tener esperanza de comprender. Debemos *ver* —al menos en principio— que cada paso del argumento puede reducirse a algo simple y obvio. La verdad matemática no es un dogma terriblemente complicado cuya validez está más allá de nuestra comprensión. Es algo construido a partir de ingredientes simples y obvios y, cuando los comprendemos, su verdad es evidente y reconocible para todos.

A mi modo de ver, esta es una reducción al absurdo tan patente como la que se pueda conseguir por otro medio, aunque esté lejos de ser una demostración matemática auténtica. El mensaje debería estar claro. La verdad matemática *no* es algo que adivinemos simplemente utilizando un algoritmo. Creo, también, que nuestra *conciencia* es un ingrediente fundamental en nuestra comprensión de la verdad matemática. Debemos "ver" la verdad de un argumento matemático para estar convencidos de su validez. Esta "visión" es la esencia misma de la conciencia. Debe estar presente *donde quiera* que percibimos directamente la verdad matemática. Cuando nos convencemos de la validez del teorema de Gödel no sólo lo "vemos" sino que al hacerlo revelamos la naturaleza no algorítmica del propio proceso de la "visión".

INSPIRACIÓN, PERSPICACIA Y ORIGINALIDAD

Intentaré hacer algunos comentarios sobre estos ocasionales soplos de intuición que conocemos como inspiración. ¿Se trata de ideas e imágenes que proceden misteriosamente de la mente inconsciente, o son producto de la propia conciencia? Podríamos citar muchos ejemplos en los que grandes pensadores han documentado tales experiencias. Como matemático me interesa especialmente el pensamiento inspirado y original de mis colegas, pero imagino que hay mucho en común entre las matemáticas y las otras ciencias y artes. Para un informe al respecto remito al lector a *The psychology of Invention in the Mathematical Field,* texto ya clásico del distinguido matemático francés Jacques Hadamard, quien cita numerosas experiencias de inspiración descritas por matemáticos de primera fila, entre otras personas. Una de las más conocidas la proporciona Henri Poincaré. Poincaré describe, en primer lugar, cómo tenía periodos intensivos de esfuerzo consciente y deliberado en su área de investigación que llamó funciones Fuchsianas, pero había llegado a un punto muerto. Entonces:

> ...Dejé Caen, en donde vivía, para participar en una excursión geológica organizada por la Escuela de Minas. Las peripecias del viaje me hicieron olvidar mi trabajo matemático. Al llegar a Coutances abordamos un autobús para ir a algún lugar. En ese momento, cuando puse mi pie en el estribo, me vino la idea, sin que nada en mis pensamiento anteriores pareciera haber preparado el camino para ello, de que las transformaciones que había utilizado para definir las funciones Fuchsianas eran idénticas a las de la geometría no euclidiana. No verifiqué la idea; no hubiera tenido

tiempo, ya que en cuanto tomé asiento en el autobús continué una conversación ya comenzada, pero tenía la certidumbre absoluta. A mi vuelta a Caen, y para quedarme tranquilo, verifiqué detenidamente el resultado.

Lo sorprendente de este ejemplo (y otros muchos citados por Hadamard) es que esta idea complicada y profunda llegó a Poincaré aparentemente en un soplo, mientras su pensamiento consciente parecía estar en otra parte, y que iba acompañada por esa sensación de estar en lo cierto como, de hecho, se demostró en los cálculos posteriores. Debería quedar claro que la misma idea no sería en absoluto fácil de explicar con palabras. Imagino que para dar una idea apropiada a personas ya expertas él habría necesitado un seminario de una hora o más. Evidentemente sólo pudo entrar de golpe en la conciencia de Poincaré, completamente formada, debido a las muchas horas previas de actividad consciente y deliberada que le familiarizó con diversos aspectos del problema. Pero, en cierto sentido, la idea que tuvo Poincaré mientras subía al autobús era una idea "simple", capaz de ser completamente captada en un instante.

Aún más notable era la convicción que tuvo Poincaré de la verdad de la idea, de modo que verificarla posteriormente y en detalle parecía casi superfluo.

Quizá debería tratar de relacionar esta experiencia con otras mías que podrían ser en cierto modo comparables. En realidad no puedo recordar ninguna ocasión en que me haya caído del cielo una idea como parece haber ocurrido en el ejemplo de Poincaré (o como en muchos otros ejemplos que se mencionan de auténtica inspiración). En mi caso parece necesario que *esté* conscientemente pensando en el problema, aunque tal vez en el fondo de mi mente. También puede ser que esté ocupado en alguna otra actividad más bien relajante; afeitarse sería un buen ejemplo. Probablemente esté empezando a pensar en un problema que había dejado de lado durante un rato. Serían ciertamente necesarias las muchas horas difíciles de actividad consciente deliberada, y a veces me llevaría un rato volver a tomar el hilo del problema. Pero la experiencia de una idea que viene "en un soplo" en tales circunstancias —junto con una fuerte sensación de convicción en su validez— no me es desconocida.

Quizá valga la pena relatar un ejemplo concreto de esto, al cual se añade un curioso interés adicional. En el otoño de 1964 había estado ocupado en el problema de las singularidades de los agujeros negros. Oppenheimer y Snyder habían demostrado, en 1939, que un colapso exactamente *esférico* de una estrella de gran masa conduciría a una singularidad central en el espacio-tiempo, en la que la teoría clásica de la relatividad general se extiende más allá de sus límites *(cfr.* capítulo VII). Muchos pensaban que esta desagradable conclusión podría evitarse si se eliminara su (poco razonable) hipótesis de simetría esférica *exacta.* En el caso esférico toda la materia se dirige hacia un mismo punto central en donde, tal vez no muy inesperadamente si se tiene en cuenta esta simetría, se produce una singularidad de densidad infinita. Parecía que no era descabellado suponer que *sin* dicha simetría la materia llegaría a la región central de una forma más desordenada y no aparecería ninguna singularidad de densidad infinita. Quizá, incluso, la materia se desplegaría de nuevo para comportarse como aquel agujero negro idealizado por Oppenheimer y Snyder.[3]

Mis propias ideas se nacían eco del renovado interés en el problema del agujero negro que se derivaba del muy reciente descubrimiento de los cuásares (a comienzos de los años sesenta). La

[3] El término "agujero negro" se hizo de uso común sólo mucho más tarde, hacia 1968 (básicamente por vía de las ideas proféticas del físico estadounidense John A. Wheeler).

naturaleza física de estos objetos astronómicos, sorprendentemente brillantes y lejanos, había llevado a mucha gente a especular que en sus centros podría residir algo semejante a los agujeros negros de Oppenheimer y Snyder. Por otro lado muchos pensaban que la hipótesis de simetría esférica de Oppenheimer-Snyder podría proporcionar una imagen completamente errónea. No obstante, tuve la idea (a partir de la experiencia de un trabajo que había hecho en otro contexto) de que debería demostrar un teorema matemático exacto que aseverara la *inevitabilidad* de las singularidades espacio-temporales (según la teoría estándar de la relatividad general), y por lo tanto reivindicar la imagen del agujero negro con tal de que el colapso hubiera alcanzado un "punto de no retorno". No conocía ningún criterio matemáticamente definible para un "punto de no retorno" (sin utilizar simetría esférica) ni mucho menos ningún enunciado o demostración de un teorema apropiado. Un colega (Ivor Robinson) me visitaba procedente de Estados Unidos y habíamos empezado una animada conversación sobre un tema muy diferente mientras paseábamos por la calle hacia mi despacho en el Birbeck College de Londres. La conversación se interrumpió al cruzar una calle lateral y se reanudó al otro lado. Por lo visto, durante esos breves instantes, se me ocurrió una idea, pero al reanudarse la conversación se borró de mi mente.

Ese mismo día, después de que mi colega hubo partido, volví a mi despacho. Recuerdo que tenía una extraña e inexplicable sensación de júbilo. Empecé a repasar en mi mente todas las cosas que me habían sucedido durante el día, intentando encontrar qué había causado mi júbilo. Después de eliminar muchas posibilidades me vino finalmente a la mente la idea que había tenido al cruzar la calle, la cual me había proporcionado una alegría momentánea dándome la solución al problema que me había estado dando vueltas en la cabeza. Al parecer ése era el criterio necesario —que posteriormente bauticé como "superficie atrapada"— y entonces no me llevó mucho tiempo dar con las líneas generales de la demostración del teorema que había estado buscando (Penrose, 1965). Aún así, pasó algún tiempo antes de que pudiera formular la demostración de un modo completamente riguroso, pero la idea que había tenido mientras cruzaba la calle había sido la clave. (A veces me pregunto qué hubiera pasado si hubiera tenido otra experiencia gozosa durante ese día. Acaso no hubiera recordado nunca la idea de la superficie atrapada.)

Esta anécdota me lleva a otro tema concerniente a la inspiración y la intuición, a saber: que los criterios *estéticos* son enormemente válidos al formar nuestros juicios. En las artes podría decir que los criterios estéticos son los supremos. La estética en las artes es una tema complejo y los filósofos han dedicado vidas enteras a su estudio. Podría argumentarse que en matemáticas y en las ciencias, tales criterios son meramente secundarios, siendo supremo el criterio de *verdad,* sin embargo, *parece* imposible separar uno de otro cuando consideramos los temas de la inspiración y la intuición. Mi impresión es que la fuerte convicción de *invalidez* de un soplo de inspiración (no ciento por ciento confiable añadiría, pero al menos mucho más confiable que el simple azar) está ligado muy estrechamente con sus cualidades estéticas. Una idea bella tiene mucha mayor probabilidad de ser correcta que una idea fea. Ésa ha sido al menos mi experiencia, y otros han expresado sentimientos similares *(cfr.* Chandrasekhar, 1987). Por ejemplo Hadamard (1945) escribe:

> ...es evidente que sin la *voluntad* de encontrar no puede tener lugar ningún descubrimiento o invención importante. Pero con Poincaré vemos algo más: el sentido de belleza desempeña su función como *medio* indispensable para encontrar. Hemos llegado a una conclusión doble: que la invención es elección, que esta elección está gobernada imperiosamente por el sentido de la belleza científica.

También Dirac (1982), por ejemplo, afirma abiertamente que fue su *agudo sentido de la belleza* el que lo hizo capaz de descubrir su ecuación para el electrón (la "ecuación de Dirac" mencionada), mientras que otros la había buscado en vano. Yo puedo ciertamente responder de la importancia de las cualidades estéticas en mi propio trabajo, tanto en lo que se refiere a la "convicción" que se sentirá con ideas que podrían calificarse posiblemente como "inspiradas" como en lo referente a las conjeturas más "rutinarias" que hemos de hacer continuamente conforme sentimos que nos acercamos al objetivo deseado. He escrito en otro lugar sobre este tema, en particular en relación con el descubrimiento de las teselas aperiódicas descritas en las figs. X.3 y IV.11. Indudablemente fueron las cualidades estéticas de la primera de estas pautas de teselación —no sólo su apariencia visual, sino también sus intrigantes posibilidades matemáticas— las que habían permitido que me viniera la intuición (tal vez en un "soplo" pero con sólo 60% de certidumbre) de que su disposición podría ser forzada mediante reglas adecuadas de ajuste (esto es, ensamblaje en anzuelo). Dentro de un momento veremos más sobre estas pautas de teselación. *(cfr.* Penrose, 1974.)

Me parece evidente que la importancia de los criterios estéticos no sólo se aplica a los juicios de inspiración instantáneos sino también a los juicios mucho más frecuentes que hacemos en el trabajo matemático (o científico). Los argumentos rigurosos constituyen normalmente el *último* paso. Antes de ello, tenemos que hacer muchas conjeturas y, para éstas, las convicciones estéticas son de enorme importancia, siempre limitadas por el razonamiento lógico y los hechos conocidos.

Son estos juicios los que considero la impronta del pensamiento consciente. Mi conjetura es que, incluso para el repentino golpe de intuición de la mente inconsciente, la *conciencia* es el arbitro y de no "sonar verdadera", la idea será rápidamente rechazada y olvidada. (Curiosamente, yo *olvidé* mi superficie atrapada, pero esto no ocurre en el nivel que entiendo ahora. La idea irrumpió en la conciencia durante el tiempo suficiente para dejar una impresión duradera.) El rechazo "estético" al que me refiero podría ser tal, supongo, que prohibiera que las ideas poco atractivas alcanzaran cualquier nivel apreciablemente permanente de la conciencia.

¿Cuál es entonces mi opinión sobre el papel del *inconsciente* en el pensamiento inspirado? Admito que estas cuestiones no son tan claras como me gustaría. Esta es un área en la que el inconsciente parece desempeñar un papel esencial, y debo coincidir en la opinión de que los procesos inconscientes son importantes. Debo conceder, asimismo, que no puede tratarse simplemente de que la mente consciente genere ideas aleatoriamente. Debe haber un proceso de selección enormemente poderoso que permite que la mente consciente sea perturbada sólo por ideas que "tienen alguna posibilidad". En mi opinión, estos criterios de selección —principalmente los "estéticos" de alguna especie— han sido ya fuertemente influidos por los *desiderata* conscientes (como la sensación de fealdad que acompaña a las ideas matemáticas incompatibles con los principios generales ya establecidos).

En relación con esto debería plantearse la cuestión de qué constituye la auténtica *originalidad.* Me parece que aquí intervienen dos factores; a saber: un proceso de "propuesta" y uno de "rechazo". Pienso que la propuesta podría ser en la mayoría de los casos inconsciente y el rechazo fundamentalmente consciente. Sin un proceso de propuesta efectivo no tendríamos ideas nuevas en absoluto. Pero este procedimiento tendría poco valor por sí solo. Necesitamos un procedimiento efectivo para formar juicios, de modo que sólo sobrevivirán las ideas que tengan una oportunidad de éxito razonable. En los sueños, por ejemplo, pueden venir fácilmente a la

mente ideas insólitas, pero sólo muy raramente sobrevivirán al juicio crítico de la conciencia despierta. (En mi caso nunca me ha venido una idea científica acertada en un sueño, mientras que otros, como el químico Kekulé con su descubrimiento de la estructura del benceno, parecen haber tenido más fortuna.) En mi opinión, es el proceso (juicio) consciente de rechazo, más que el proceso inconsciente de propuesta, el que es capital en el tema de la originalidad; pero sé que muchos otros podrán sostener una opinión contraria.

Antes de dejar las cosas en este estado más bien insatisfactorio mencionaré otra característica sorprendente del pensamiento inspirado; a saber: su carácter *global.* La anécdota anterior de Poincaré era un ejemplo sorprendente puesto que la idea que vino a su mente en un momento fugaz ha dominado un área enorme del pensamiento matemático.

Quizá sea más inmediatamente accesible al lector no matemático (aunque no más comprensible, sin duda) el modo en que algunos artistas pueden tener en su mente, a la vez, la totalidad de sus creaciones Mozart nos da un ejemplo vivido y sorprendente (citado en Hadamard, 1945 p. 16):

> Cuando me siento bien y de buen humor, o cuando doy un paseo en carruaje o a pie tras una buena comida, o por la noche cuando no puedo dormir las ideas se agolpan en mi mente tan fácilmente como lo quiera. ¿De dónde y cómo vienen? Lo ignoro y nada tengo que ver con ello. Guardo en la mente las que me gustan y las tarareo; al menos eso me han dicho que hago, Una vez que tengo el tema, viene una melodía ligada con la primera de acuerdo con las necesidades de la composición: el contrapunto, la parte de cada instrumento y por último, todos los fragmentos melódicos dan lugar a la obra completa, Entonces mi espíritu se inflama de Inspiración, La obra crece; sigo desarrollándola, concibiéndola cada vez con más claridad hasta que la tengo acabada por larga que pueda ser. Entonces mi mente la atrapa de la misma forma en que mi ojo atrapa de una mirada una imagen bella o la hermosura de la juventud. No viene a mí poco a toco, con las diversas partes trabajadas en detalle a medida que se van haciendo, sino que es su totalidad como me deja oírlas mí imaginación.

Me parece que esto concuerda con un esquema de propuesta-rechazo. La propuesta parece ser inconsciente ("no tengo nada que ver con ello") aunque, sin duda, fuertemente selectiva, mientras que el rechazo es el arbitro consciente del gusto ("guardo las que me gustan..."). La globalidad del pensamiento inspirado es particularmente notable en la cita de Mozart ("no me viene poco a poco... sino es su totalidad") y también en Poincaré ("no verifiqué la idea; no hubiera tenido tiempo"). Además, sostendré que una notable globalidad está ya presente en general en nuestro pensamiento consciente. Volveré a esta cuestión en un momento.

LA NO VERBALIDAD DEL PENSAMIENTO

Uno de los principales señalamientos que hace Hadamard en su estudio sobre el pensamiento creativo es su impresionante refutación de la tesis, todavía tan socorrida, de que la verbalización es necesaria para el pensamiento. Difícilmente podríamos hacer algo mejor que repetir una cita de una carta que recibió de Albert Einstein a propósito de esta cuestión:

Las palabras o el lenguaje, ya sea escrito o hablado, no parecen desempeñar ningún papel en mi mecanismo de pensamiento. Las entidades físicas que parecen servir como elementos del

pensamiento son ciertos signos e imágenes más o menos claros que pueden reproducirse y combinarse "voluntariamente"... Los elementos antes mencionados son, en mi caso, de tipo visual y muscular Las palabras u otros signos convencionales tienen que buscarse laboriosamente sólo en una segunda etapa, cuando el citado juego asociativo está suficientemente establecido y puede ser reproducido a voluntad.

También merece ser citado el eminente genetista Francis Galton:

> Resulta una seria desventaja para mí al escribir, y aun más el explicar oralmente, el hecho de que no pienso tan fácilmente en palabras como en otras formas. Sucede a menudo que después de hacer un trabajo duro, y habiendo llegado a resultados que son perfectamente claros y satisfactorios para mí, cuando trato de expresarlos en el lenguaje conveniente siento que debo empezar a colocarme en un plano intelectual muy distinto. Tengo que traducir mis ideas en un lenguaje que no corre muy a la par con ellas. Por ello gasto mucho tiempo en buscar palabras y frases apropiadas, y soy consciente, cuando tengo que hablar de improviso, de ser a menudo muy oscuro por simple torpeza verbal y no por falta de claridad de percepción. Este es uno de los pequeños fastidios de mi vida.

El propio Hadamard escribe también:

> Insisto en que las palabras están totalmente ausentes de mi mente cuando realmente pienso, y mi caso está en la misma línea de Galton en el sentido de que, incluso después de leer o escuchar una pregunta, todas las palabras desaparecen en el preciso instante en que empiezo a pensar sobre ello; y coincido plenamente con Schopenhauer cuando escribe "las ideas mueren en el momento en que se encarnan en palabras".

Cito estos ejemplos porque se ajustan mucho a mis propios modos de pensamiento. Casi todo mi pensamiento matemático se construye visualmente y en términos de conceptos no verbales, aunque con frecuencia las ideas vayan acompañadas de algún comentario verbal insustancial y casi inútil, tal como "esto va con aquello y aquello va con eso otro". (Pudiera utilizar a veces palabras para simples inferencias lógicas.) También he experimentado frecuentemente por mí mismo las dificultades que tenían estos pensadores para traducir sus pensamientos en palabras. Con frecuencia la razón es simplemente que no hay palabras disponibles para expresar los conceptos requeridos. De hecho, yo calculo a menudo con diagramas especialmente diseñados que constituyen una especie de taquigrafía para ciertos tipos de expresiones algebraicas *(cfr.* Penrose y Rindler, 1984, pp. 424, 434). Sería así un proceso muy molesto tener que traducir estos diagramas en palabras, y esto es algo que sólo hago como último recurso si se hace necesario para dar una explicación detallada a los demás. En relación con esto he notado, en ocasiones, que si me he estado concentrando intensamente en matemáticas durante un rato y alguien inicia repentinamente una conversación, entonces me encuentro casi incapaz de hablar durante varios segundos. Esto no quiere decir que no piense a veces con palabras, sino que simplemente encuentro las palabras casi inútiles para el pensamiento *matemático.* Otros tipos de pensamiento, quizá tales como el *filosofar* parecen mucho más adecuados para la expresión verbal. Quizá sea esta 1a razón de que muchos filósofos parecen ser de la opinión de que el lenguaje es esencial para el pensamiento inteligente o consciente. Sin duda personas diferentes piensan de modos muy diferentes, como por cierto ha ocurrido en mi propia experiencia, incluso

entre matemáticos. La polaridad analítico-geométrico parece ser la principal en el pensamiento matemático. Es interesante que Hadamard se considere analítico pese a que utilizaba imágenes visuales más que verbales para su pensamiento matemático. En cuanto a mí, estoy casi en el extremo geométrico de las cosas, pero el espectro entre los matemáticos es generalmente muy amplio.

Una vez que se acepta que mucho del pensamiento consciente puede ser de carácter no verbal —y, en mi opinión, esta conclusión se desprende inevitablemente de las anteriores consideraciones— entonces quizá el lector no encuentre tan difícil creer que tal pensamiento pudiera tener también un ingrediente no algorítmico.

Recuérdese que el capítulo IX mencioné el punto de vista, expresado con frecuencia, de que sólo la mitad del cerebro que es capaz del habla (la mitad izquierda, en la inmensa mayoría) sería también capaz de la conciencia. En vista de lo dicho, está claro para el lector por qué encuentro este punto de vista de) todo inaceptable. Yo no sé si, en general, los matemáticos tenderán a utilizar más un lado del cerebro; pero no puede haber duda del alto grado de conciencia que se necesita para el auténtico pensamiento matemático. Mientras que el pensamiento analítico parece ser principalmente competencia del lado izquierdo del cerebro, se argumenta a menudo que el pensamiento geométrico reside en el lado *derecho,* de modo que es una conjetura muy razonable el que mucha de la actividad matemática *consciente tenga lugar* en el lado derecho.

¿CONCIENCIA ANIMAL?

Antes de dejar el tema de la importancia de la verbalización para la conciencia me ocuparé de la cuestión, antes planteada brevemente, de si los animales pueden ser conscientes. Me parece que a veces la gente se basa en la incapacidad de los animales para hablar, como un argumento en contra de que puedan tener conciencia apreciable alguna y, consecuentemente, en contra de que tengan "derechos". El lector se dará perfecta cuenta de que considero ésta una línea de razonamiento insostenible, ya que mucho pensamiento complejo consciente (p.ej. el matemático) puede de llevarse a cabo sin verbalización. También se argumenta a veces que el lado derecho del cerebro tiene una conciencia pequeña, como la un chimpancé, debido también a su falta de capacidad verbal *(cfr.* LeDoux 1985, pp. 197-216).

Existe una importante controversia sobre si los chimpancés y los gorilas son capaces de auténtica verbalización cuando se les permite utilizar un *lenguaje de signos* en lugar de la forma humana normal de hablar (lo que ellos no pueden hacer debido a la falta de cuerdas vocales adecuadas). (Véanse diversos artículos en Blakemore y Greenfield, 1987.) Parece claro, pese a la controversia, que estos animales son capaces de comunicar, al menos hasta cierto grado elemental, por estos medios. A mi modo de ver, resulta cicatero por parte de algunas personas el no admitir que a eso pueda llamársele "verbalización". Quizá negándoles a los monos la entrada en el club de los verbalizadores, algunos esperen excluirlos del club de los seres conscientes.

Dejando aparte la cuestión del habla, existen testimonios suficientes de que los chimpancés son capaces de genuina *inspiración.* Konrad Lorenz (1972) describe a un chimpancé en una habitación en la que había un plátano suspendido del techo apenas fuera de su alcance, y una caja en otro lado de la habitación:

La cuestión no lo dejaba en paz y volvía una y otra vez al problema. Entonces, súbitamente —y no hay otra forma de describirlo— su cara antes deprimida "se iluminó". Sus ojos se movían desde el plátano hacia el espacio vacío que había bajo la caja, y desde éste a la caja, luego volvía al espacio y allí al plátano. A continuación dio un grito de alegría y luego dio una voltereta sobre la caja de pura animación. Completamente seguro de su éxito, empujó la caja hasta ponerla debajo del plátano. Nadie que lo observara podría dudar de la existencia de una auténtica experiencia de discernimiento en los simios antropoides.

Nótese que, como en la experiencia de Poincaré cuando abordó el ómnibus, el chimpancé estaba "completamente seguro de su éxito" antes de haber verificado su idea. Si estoy en lo correcto en cuanto a que tales juicios requieren de conciencia, entonces ésta es una evidencia de que animales no humanos pueden ser conscientes.

Una cuestión interesante se plantea en relación con los delfines (y las ballenas). Puede señalarse que los cerebros de los delfines son tan grandes (o mayores) que los nuestros, y los delfines pueden enviarse también complejas señales sonoras de uno a otro. Muy bien pudiera ser que sus grandes cerebros fueran necesarios para algún propósito distinto de la "inteligencia" a escala humana o casi humana. Además, debido a su falta de manos prensiles, no son capaces de construir una "civilización" de tipo de las que podemos apreciar y aunque, por la misma razón no puedan escribir libros, podrían a veces filosofar y meditar sobre el sentido de la vida y por qué están "allí" ¿Podrían transmitir a veces sus sensaciones de "conciencia" por vía de sus complejas señales sonoras submarinas? No sé de ninguna investigación que indique si ellos utilizan un lado particular de sus cerebros para "verbalizar" y comunicarse con otros. Respecto a las operaciones de "escisión cerebral" que se han realizado en humanos, con sus enigmáticas implicaciones sobre la continuidad del "yo", debería subrayarse que no todo el cerebro del delfín duerme [4] simultáneamente sino que los hemisferios cerebrales duermen alternadamente. Sería instructivo poder preguntarles cómo "sienten" la continuidad de la conciencia.

CONTACTO CON EL MUNDO DE PLATÓN

He mencionado que parece haber muchas diferentes maneras de pensar para personas diferentes e incluso diferentes maneras de pensar sobre matemáticas entre los matemáticos. Recuerdo que cuando estaba a punto de entrar en la universidad para hacer mis estudios esperaba encontrar que mis colegas matemáticos pensaran más o menos como yo lo hacía. Mi experiencia escolar había sido que mis compañeros de clase parecían pensar de una forma bastante diferente, lo que había encontrado bastante desconcertante. "Ahora", me había dicho entusiasmado, "encontraré colegas con los que pueda comunicarme libremente. Algunos pensarán con más eficacia que yo, y algunos con menos; pero todos ellos estarán en mi misma frecuencia de pensamiento". ¡Qué equivocado estaba! Creo que encontré más diferencias en cuanto a modos de pensar de las que había experimentado hasta entonces. Mi propio pensamiento era mucho más geométrico y menos analítico que el de los demás, pero había muchas otras diferencias entre las formas de pensar de mis diversos colegas. Siempre tuve especial dificultad para comprender la descripción verbal de una fórmula, mientras que muchos de mis colegas parecían no experimentar tal dificultad.

[4] Me parece que el hecho de que los animales necesiten un periodo de sueño en el que parecen a veces soñar (como se nota a menudo en los perros) arguye que pueden poseer conciencia. En efecto, algo de conciencia parece ser un ingrediente importante en la distinción entre el dormir y el soñar.

Una experiencia común, cuando algún colega trataba de explicarme algún aspecto de las matemáticas, era que a menudo lo escuchaba atentamente pero sin comprender prácticamente nada de las conexiones lógicas entre un conjunto de palabras y el siguiente. Sin embargo, en mi mente se formaba cierta imagen conjetural para las ideas que él trataba de mostrarme —formada completamente en mis propios términos y al parecer con poca conexión con las imágenes mentales que habían sido la base de la propia comprensión de mi colega— y yo respondía. Para mi asombro, por lo general mis comentarios se aceptaban como apropiados, y la conversación continuaba así. Estaba claro, al fin, que había tenido lugar alguna comunicación positiva. Pero las oraciones que pronunciábamos cada uno, sólo muy rara vez parecían comprenderse. En mis años subsecuentes de matemático (o físico-matemático) profesional he encontrado que este fenómeno no es menos cierto de lo que era en mi época de estudiante. Quizá, a medida que ha aumentado mi experiencia matemática, he imaginado mejor lo que los otros quieren dar a entender con sus explicaciones, y quizá soy un poco mejor para tener en cuenta los otros modos de pensar cuando yo mismo me explico las cosas; pero, en esencia, nada ha cambiado.

Con frecuencia me he devanado la cabeza pensando en cómo es posible cualquier comunicación según este extraño procedimiento, pero ahora me gustaría aventurar una especie de explicación, pues pienso que pudiera ser de profunda importancia para las otras cuestiones que he abordado. La idea es que al transmitir las matemáticas *no* simplemente comunicamos *hechos.* Para que pueda comunicarse de una persona a otra una cadena de hechos (contingentes) es necesario que la primera los enuncie cuidadosamente y que la segunda los pueda aceptar individualmente. Pero en las matemáticas el contenido *fáctico* es pequeño. Los enunciados matemáticos son verdades necesarias (o, como por el contrario, falsedades necesarias) e incluso si el enunciado del primer matemático representa meramente un tanteo de verdad necesaria será esa misma verdad la que se transmitirá al segundo matemático con tal de que éste la haya comprendido de forma adecuada. Las imágenes mentales del segundo pueden diferir de las del primero igual que pueden diferir sus descripciones verbales, pero la idea matemática importante habrá pasado de uno a otro.

Ahora bien, este tipo de comunicación no sería en absoluto posible si no fuera por el hecho de que las verdades matemáticas *interesantes* o *profundas* están mínimamente distribuidas entre las verdades matemáticas en general. Si la verdad que se va a transmitir fuera, pongamos por caso, el enunciado sin interés 4897 x 512 = 2507264, entonces el segundo tendrá que haber comprendido realmente del primero para que se transmita el enunciado exacto. Pero para un enunciado matemático interesante podemos darnos cuenta a menudo del concepto que se intenta transmitir incluso si se ha proporcionado una descripción muy imprecisa del mismo.

Podría parecer que hay una paradoja en esto, ya que la matemática es un tema en donde reina la precisión. En realidad, en las presentaciones escritas se tiene mucho cuidado para asegurarse de que los diversos enunciados son a la vez precisos y completos. Sin embargo, para transmitir una idea matemática (normalmente en descripciones verbales) semejante precisión puede tener a veces un efecto inhibidor a primera vista, y puede resultar necesaria una forma de comunicación más vaga y descriptiva. Una vez que la idea ha sido captada en esencia, entonces pueden examinarse los detalles.

¿Cómo es posible que se puedan comunicar las ideas matemáticas de esta forma? Imagino que siempre que la mente percibe una idea matemática toma contacto con el mundo platónico de los conceptos matemáticos. (Recordemos que, según el punto de vista platónico, las ideas matemáticas tienen una existencia propia, y habitan en un mundo ideal que sólo es accesible por

la vía del intelecto.) Cuando "vemos" una verdad matemática, nuestra conciencia irrumpe en este mundo de ideas y toma contacto directo ("accesible por vía del intelecto"). He descrito esta "visión" en relación con el teorema de Gödel pero es la esencia de la comprensión matemática. Cuando los matemáticos se comunican, esto se hace posible porque cada uno tiene un *camino directo* a la verdad, estando su conciencia en disposición de percibir directamente las verdades matemáticas a través de estos procesos de "visión". (De hecho, este acto de percepción va acompañado a menudo de expresiones como "¡ah, ya veo!") Puesto que cada uno puede tomar contacto directamente con el mundo platónico, pueden comunicarse entre ellos mucho más fácilmente de lo que pudieran esperar. Las imágenes mentales que tienen cuando hacen este contacto platónico podrían ser bastante diferentes para cada uno, pero la comunicación es posible porque todos están en contacto directo con el *mismo* mundo platónico externamente existente.

Según esta idea, la mente es siempre capaz de este contacto directo. Pero sólo un poco puede trasvasar cada vez. El descubrimiento matemático consiste en un ensanchamiento del área de contacto. Debido al hecho de que las verdades matemáticas son verdades necesarias, ninguna "información" real, en el sentido técnico, pasa al descubridor. Toda la información estaba allí todo el tiempo, era sólo cuestión de atar los cabos y "ver" la respuesta. Esto coincide prácticamente con la propia idea de Platón de que el descubrimiento (digamos matemático) es sólo una forma de *recuerdo*. En realidad, me he sorprendido a menudo del parecido entre el hecho de no poder recordar el nombre de alguien y el de no ser capaz de encontrar el concepto matemático correcto. En ambos casos, el concepto buscado en cierto sentido está *ya presente* en la mente, aunque esta es una forma menos usual de hablar en el caso de una idea matemática no descubierta.

Para que esta manera de ver las cosas sea útil, en el caso de la comunicación matemática, debemos imaginar que las ideas matemáticas interesantes y profundas tienen de algún modo una existencia más fuerte que las poco interesantes o triviales. Esto tendrá importancia en relación con las consideraciones especulativas de la próxima sección.

UNA VISIÓN DE LA REALIDAD FÍSICA

Cualquier punto de vista acerca de la aparición de la conciencia dentro del universo de la realidad física debe ocuparse, al menos implícitamente, de la cuestión de la propia realidad física.

El punto de vista de la IA fuerte, por ejemplo, sostiene que la "mente" es la encarnación de un algoritmo suficientemente complejo activado por algunos objetos del mundo físico. Se supone que no importa cuáles sean estos objetos reales. Igual valdrán señales nerviosas que corrientes eléctricas en cables, engranajes, poleas o tuberías. Todo lo que importa es el propio algoritmo. Pero para que un algoritmo "exista" independientemente de cualquier encarnación física particular parece ser esencial un punto de vista platónico. Sería difícil para un defensor de la IA fuerte aceptar que "los conceptos matemáticos existen sólo en la mente", ya que esto sería un círculo vicioso que requiere la existencia previa de la mente para que existan algoritmos y la existencia de algoritmos para que haya mentes. Dicho defensor podría tratar de adoptar la postura de que los algoritmos pueden existir como marcas en un trozo de papel, o direcciones de magnetización en un bloque de hierro, o desplazamientos de cargas en la memoria de una computadora. Pero tales configuraciones de material no constituyen por sí mismas un algoritmo.

Para convertirse en algoritmos necesitan una *interpretación,* es decir, debe ser posible *decodificar* las configuraciones; y ello dependerá del lenguaje en el que estén escritos los algoritmos. Una vez más, parece que se necesita una mente preexistente para "comprender" el lenguaje, con lo que volvemos a donde estábamos. Aceptando entonces que los algoritmos residen en el mundo platónico y, por lo tanto, que en *dicho* mundo, según la idea de la IA fuerte, es donde deben encontrarse las mentes, tenemos que afrontar ahora la cuestión de cómo pueden relacionarse el mundo físico y el mundo platónico. Ésta es, a mi parecer, la versión de la IA fuerte del problema mente-cuerpo.

Mi propio punto de vista es diferente de éste, puesto que yo creo que las mentes (conscientes) *no* son entidades algorítmicas. Sin embargo, no deja de desconcertarme descubrir que existen muchos puntos e entre el punto de vista de la IA fuerte y el mío. He señalado que creo que la conciencia está estrechamente asociada con la sensación de verdades necesarias —y por consiguiente con la consecución de un contacto directo con el mundo platónico de los conceptos matemáticos. Esto no es un procedimiento algorítmico y no es el algoritmo que pudiera habitar en ese mundo el que nos interesa especialmente, pero de nuevo se ve que el problema mente-cuerpo, según esta idea, está íntimamente ligado a la cuestión de cómo se relaciona el mundo platónico con el mundo de los objetos físicos.

Hemos visto en los capítulos V y VI cómo el mundo físico parece ajustarse de un modo notable a algunos esquemas matemáticos precisos (las teorías SUPREMAS). Se ha insistido a menudo en lo extraordinaria que es esta precisión *(cfr.* especialmente Wigner, 1960). Me resulta difícil creer, como algunos tratan de sostener, que semejantes teorías SUPREMAS pueden haber surgido simplemente por alguna selección natural de ideas que dejase sólo las buenas como supervivientes. Las buenas lo son *demasiado* como para ser simplemente las supervivientes de ideas que hayan aparecido de modo aleatorio. En lugar de ello, debe haber alguna poderosa razón para el acuerdo entre las matemáticas y la física, es decir, entre el mundo platónico y el mundo físico.

Al hablar del "mundo platónico" le asignamos un tipo de realidad comparable, en algún sentido a la realidad del mundo físico. Por otra parte, la realidad del propio mundo físico parece más nebulosa de lo que parecía antes de la llegada de las teorías supremas de la relatividad y la mecánica cuántica . La misma exactitud de esas teorías ha proporcionado una existencia matemática casi abstracta para la verdadera realidad física. ¿Es esto de alguna forma una paradoja? ¿Cómo puede la realidad concreta hacerse abstracta y matemática? Este es tal vez el otro lado de la pregunta de cómo los conceptos matemáticos abstractos pueden conseguir una realidad casi concreta en el mundo platónico. ¿Quizá, en cierto sentido, los dos mundos son realmente el *mismo? (Cfr.* Wigner, 1960; Penrose, 1979a; Barrow, 1988; también Atkins, 1987.)

Aunque siento una fuerte simpatía por la idea de identificar estos dos mundos, debe haber algo más que eso. Como he mencionado en el capítulo III, y en una sección precedente de este mismo capítulo, algunas verdades matemáticas parecen tener una realidad platónica más fuerte (¿"más profunda", "más interesante", "más fructífera"?) que otras. Tales verdades serían las que se identificarían más estrechamente con el funcionamiento de la realidad física. (El sistema de los números complejos *[cfr. capítulo III]* sería un caso concreto por ser éstos los ingredientes fundamentales de la mecánica cuántica, las amplitudes de probabilidad.) Con semejante identificación podría ser más comprensible cómo la "mente" parecía manifestar alguna curiosa conexión entre el mundo físico y el mundo platónico de las matemáticas. Recordemos que, corno

se escribió en el capítulo IV, existen muchas panes del mundo matemático —algunas de sus partes más profundas y mas interesantes, por lo demás— que no tienen carácter algorítmico. Parece probable, por lo tanto, sobre la base del punto de vista que trato de exponer, que la acción no algorítmica debería tener un papel de importancia muy considerable dentro del mundo físico. Me atrevo a afirmar que este papel está íntimamente ligado con el propio concepto de "mente".

DETERMINISMO Y DETERMINISMO FUERTE

Hasta aquí poco he dicho sobre la cuestión del "libre albedrío", que normalmente se considera el tema fundamental del aspecto *activo* del problema mente-cuerpo. En lugar de ello me he concentrado en mi proposición de que existe un aspecto *no algorítmico* en el papel de la acción consciente. Normalmente, el tema del libre albedrío se estudia en relación con el determinismo en física. Recordemos que en la mayoría de nuestras teorías supremas existe un determinismo estricto, en el sentido de que si se conoce el estado del sistema en un instante cualquiera,[5] entonces está completamente fijado para todos los instantes posteriores (o también anteriores) por las ecuaciones de la teoría. Por consiguiente, parece que no hay lugar para el libre albedrío puesto que el comportamiento futuro de un sistema parece estar totalmente determinado por las leyes físicas. Incluso la parte U de la mecánica cuántica tiene este carácter completamente determinista. Sin embargo, la parte R del "salto cuántico" no es determinista e introduce un elemento completamente aleatorio en la evolución temporal. Con anterioridad, diversas personas aventuraron ya la responsabilidad de que pudiera haber un papel para el libre albedrío, y que la acción de la conciencia tuviera quizá algún efecto directo sobre el modo en que podría saltar un sistema cuántico individual. Pero si R es *realmente* aleatorio, entonces no sirve de gran ayuda si deseamos hacer algo positivo con nuestras voluntades.

Mi propio punto de vista, aunque no esté muy bien formulado a este respecto, sería el de que algún nuevo procedimiento (GQC; *cfr.* capítulo VIII) toma el mando a partir de la línea divisoria cuántico-clásico e interpola entre U y R (que ahora se consideran como aproximaciones), y que este nuevo procedimiento contendría un elemento esencialmente *no algorítmico*. Esto implicaría que el futuro *no sería computable* a partir del presente, incluso *aunque pudiera estar determinado* por él. He tratado de ser claro al distinguir el tema de la computabilidad del tema del determinismo en mi exposición del capítulo V. Me parece bastante probable que la GQC sea una teoría determinista pero no computable.[*] (Recordemos el "modelo de juguete" no computable que se describió en el capítulo V.)

Hay quienes aceptan la idea de que ni siquiera con el determinismo clásico (o U cuántico) existe determinismo efectivo, ya que las condiciones iniciales nunca podrían conocerse suficientemente bien para que se pudiera *verdaderamente* computar el futuro. A veces, cambios muy pequeños en las condiciones iniciales pueden conducir a diferencias muy grandes en el resultado final. Esto es lo que sucede, por ejemplo, en el fenómeno conocido como "caos" en un sistema determinista (clásico), y un ejemplo de ello es la incertidumbre en la predicción del tiempo meteorológico. Sin embargo, es muy difícil creer que este tipo de incertidumbre clásica pudiera ser la que nos

---

[5] En el caso de la relatividad especial o general, léase "espacios simultáneos" o "superficies de tipo especial" en lugar de "tiempos".

[*] Puede señalarse que hay al menos una aproximación a una teoría de gravitación cuántica que parece implicar un elemento de no computabilidad (Geroch y Hartle, 1987).

permita tener (¿la ilusión de?) libre albedrío. El comportamiento futuro seguirá estando *determinado,* desde la misma gran explosión, incluso aunque seamos incapaces de calcularlo.

La misma objeción podría hacerse contra mi hipótesis de que la falta de *computabilidad* podría ser intrínseca a las leyes dinámicas —que ahora se suponen de carácter no algorítmico— antes que a nuestra falta de información acerca de las condiciones iniciales. En esta perspectiva, incluso aunque fuera no computable, el futuro seguiría estando completamente *fijado* por el pasado, de manera total desde el *big bang*. De hecho, no voy a ser tan dogmático como para insistir en que la GQC debería ser determinista pero no computable. Mi conjetura sería que la teoría buscada tendría una descripción más sutil que esa. Sólo planteo que debería contener elementos de un tipo especialmente no algorítmico.

Para cerrar esta sección me gustaría comentar una idea aún más extrema que se podría mantener sobre el tema del determinismo. Ésta es la que he denominado *determinismo fuerte* (Penrose, 1987b). Según el determinismo fuerte no se trata sólo de que el futuro esté determinado por el pasado: la *historia entera del Universo está fijada,* según algún esquema matemáticamente preciso, *para cualquier instante.* Semejante punto de vista podría resultar atractivo si nos inclinamos a identificar de alguna forma el mundo platónico con el mundo físico, puesto que el mundo platónico está fijado de una vez por todas, sin "posibilidades alternativas" para el Universo. (Me pregunto a veces si Einstein habrá tenido este esquema en la mente cuando escribió: "Lo que realmente me interesa es si Dios podía haber hecho el mundo de una forma diferente; es decir, si la necesidad de simplicidad lógica deja alguna libertad" (Carta a Ernst Strauss; véase Kuznetsov, 1977, p. 285).

Como una variante del determinismo fuerte podríamos considerar la idea de los *muchos universos* en mecánica cuántica *(cfr.* capítulo VI, p, 355). Según ésta, no sería una *simple* historia del Universo la que estaría fijada por un esquema matemáticamente preciso, sino que sería la totalidad de miríadas de miríadas de historias del Universo "posibles" las que estarían determinadas. Pese a la naturaleza desagradable (al menos para mí) de semejante esquema, y la multitud de problemas e insuficiencias que nos presenta, no podemos descartarlo como posibilidad.

Me parece que si tenemos determinismo fuerte, pero *sin* muchos universos, entonces el esquema matemático que gobierna la estructura del Universo *tendría* que ser probablemente no algorítmico.[6] De otro modo podríamos calcular en principio lo que íbamos a hacer a continuación y entonces podríamos "decidir" hacer algo muy diferente, lo que constituiría una contradicción flagrante entre el "libre albedrío" y el determinismo fuerte de la teoría. Podemos evitar esta contradicción introduciendo no computabilidad en la teoría, aunque tengo que confesar que me siento algo molesto con este tipo de solución, y espero algo mucho más sutil para las *verdaderas* reglas (no algorítmicas) que gobiernan la forma en que funciona el mundo.

[6] Existe, sin embargo, un obstáculo en el caso de un universo espacialmente infinito, ya que entonces (de modo análogo al caso de muchos universos) resulta que habría infinitas copias de uno mismo y del entorno inmediato de uno. El comportamiento futuro de cada ejemplar podría ser ligeramente diferente, y nunca estaríamos bastante seguros sobre cuál de las copias aproximadas de uno mismo modeladas en las matemáticas podríamos "ser" realmente.

EL PRICIPIO ANTRÓPICO

¿Qué importancia tiene la conciencia para el Universo en su totalidad? ¿Podría existir un universo sin habitantes conscientes? ¿Están las leyes de la física especialmente diseñadas para permitir la existencia de vida consciente? ¿Hay algo especial en nuestra localización particular en el Universo, ya sea en el espacio o en el tiempo? Este es el tipo de preguntas que pretende responder lo que se ha llegado a conocer como el *principio antrópico.*

Este principio tiene muchas formas. (Véase Barrow y Tipler, 1986.) La más aceptable de estas formas se refiere a la localización espacio-tempora1 de la vida consciente (o inteligente") en el Universo, Este es el principio antrópico *débil* Puede utilizarse el argumento para explicar por que se dan las condiciones justas para la existencia de la vida (inteligente) en la Tierra en la época presente: si no fueran las correctas entonces nosotros mismos no estaríamos aquí, sino en alguna otra parte, en alguna otra época apropiada. Este principio fue utilizado de modo muy eficaz por Brandon Carter y Robert Dicke para resolver un problema que había intrigado a los físicos durante muchos años. El problema se refería a varias relaciones numéricas sorprendentes que se han observado entre las constantes físicas (la constante gravitatoria, la masa del protón, la edad del Universo, etc.). Un aspecto enigmático de esto era que algunas de estas relaciones son válidas solamente en la época actual de la historia de la Tierra, de modo que casualmente parece que estamos viviendo un momento muy especial (con un margen de unos pocos millones de años más o menos). Esto fue finalmente explicado por Cárter y Dicke, por el hecho de que esta época coincide con la vida media de las llamadas estrellas de la secuencia principal, como es el Sol. En cualquier otra época, sigue diciendo el argumento, no existiría vida inteligente para medir las constantes físicas en cuestión, de modo que la conciencia *tenía* que darse simplemente por el hecho de que sólo existiría vida inteligente en el momento particular en que se *diera* la coincidencia.

El principio antrópico *fuerte* va más lejos. En este caso, no sólo estamos interesados en nuestra localización espacio-temporal en el Universo, sino de una infinidad de universos *posibles.* Ahora podemos sugerir respuestas a preguntas como la de por qué las constantes de la física, o las leyes de la física en general, están especialmente diseñadas para que pueda existir vida inteligente. El razonamiento sería que si las constantes o las leyes fueran diferentes, entonces no estaríamos en este Universo sino en algún otro. En mi opinión, el principio antrópico fuerte tiene un carácter algo dudoso y los teóricos tienden a invocarlo cuando no tienen ninguna teoría bastante buena para explicar los hechos observados (por ejemplo, en las teorías de la física de partículas, en las que no están explicadas las masas de las partículas y se argumenta que, si hubieran tenido valores distintos de los observados, entonces la vida presumiblemente sería imposible, etc.). El principio antrópico débil, por el contrario, me parece incuestionable siempre que seamos muy cuidadosos en la forma de utilizarlo.

Mediante la utilización del principio antrópico —ya sea en la forma fuerte o en la débil— podríamos tratar de probar que la conciencia era inevitable en virtud del hecho de que tenía que haber seres sensibles, es decir "nosotros", para observar el mundo, así que *no* necesitamos suponer, como he hecho, que la sensibilidad tenga alguna ventaja selectiva.

En mi opinión, este razonamiento es técnicamente correcto y el argumentó del principio antrópico débil *podría* proporcionar (al menos) una razón para que la conciencia exista sin que tenga que ser favorecida por la selección natural. Pese a ello, yo no puedo creer que el argumento

antrópico sea la razón *auténtica* (o la única razón) para la evolución de la conciencia. Hay suficientes pruebas procedentes de otras direcciones para convencerme de que la conciencia *tiene* una poderosa ventaja selectiva, y no creo que el principio antrópico sea necesario.

TESELACIONES Y CUASICRISTALES

Dejaré ahora las especulaciones demasiado generales de las últimas secciones y consideraré en su lugar una cuestión que, aunque todavía algo especulativa, es mucho más científica y "tangible". Esta cuestión parecerá al principio una digresión intrascendente; pero su importancia para nosotros se pondrá de manifiesto en la próxima sección.

Recordemos las pautas de teselación mostradas en la fig. IV.12. Estas pautas son notables: "casi" violan un conocido teorema matemático relativo a las redes cristalinas. El teorema establece que las únicas teorías racionales que están permitidas para una estructura cristalina son dobles, triples, cuádruples y séxtuples. Por estructura cristalina entiendo un sistema discreto de puntos que tiene una *simetría de traslación;* es decir, existe una manera de hacer deslizar la estructura sobre sí misma sin rotarla de modo que finalmente la estructura coincida consigo misma (esto es, invariante bajo ese movimiento particular) y por lo tanto tiene un *paralelogramo de períodos* (*cfr.* fig- IV.8). Ejemplo de pautas de teselación con esas simetrías de rotación se dan en la fig. X.2. Ahora bien, las pautas de la fig. IV. 12, como las mostradas en la fig. X.3 (que es esencialmente la teselación producida al ajustar las teselas de la fig. IV. 11) *casi* tienen, por el contrario, simetrías de traslación y *casi* tienen simetría *quíntuple,* en donde "casi" significa que podemos encontrar movimientos de la estructura (traslacional y rotacional respectivamente) tales que la estructura coincide consigo misma hasta cualquier grado preasignado que se desee, aunque sin llegar al ciento por ciento. No tenemos que preocuparnos ahora del significado exacto de esto. El único punto que será importante para nosotros es que si tuviéramos una sustancia en la que los átomos estuvieran dispuestos en los vértices de esta estructura, entonces parecería ser cristalina pese a que exhibiría una simetría quíntuple prohibida.

En diciembre de 1984, el físico israelí Dany Shechtman, que había estado trabajando con otros colegas en el National Bureau of Standards en

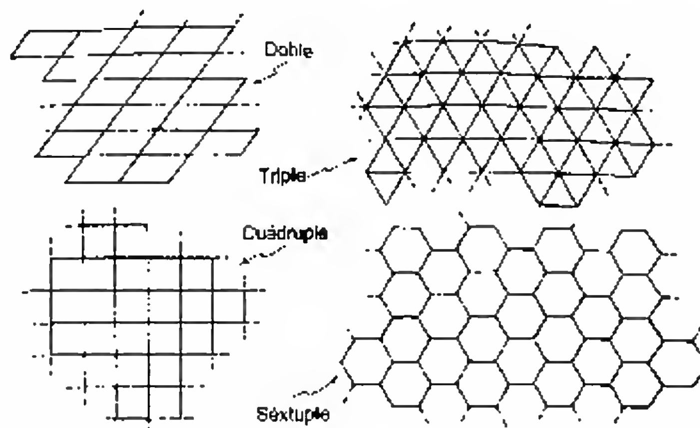


FIGURA X.2. *Teselaciones periódicas con diversas simetrías (en las que el centro de simetría, en cada caso, se considera que es el centro de la tesela).*

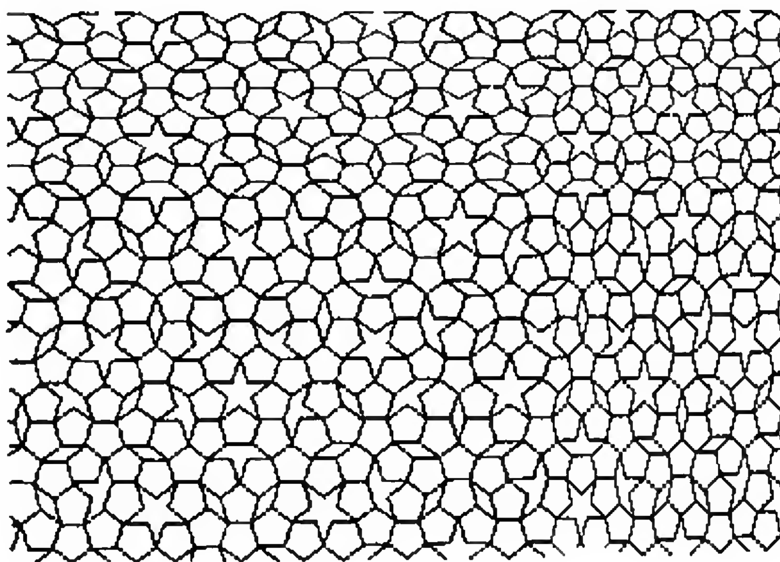


FIGURA X.3. *Teselación cuasiperiódica (básicamente la que se produce ajustando las teselas de la fig. IV. 11) con una cuasisimetría quíntuple cristalográficamente "imposible".*



FIGURA X.4. *Cuasicristal (aleación At-Li-Cu) con una simetría cristalina aparentemente imposible. (Tomado de Gayle, 1987.)*

Washington DC, anunció el descubrimiento de una fase de una aleación de aluminio-manganeso que parecía ser realmente una sustancia de tipo cristalino —ahora denominada *cuasicristal*— con simetría quíntuple. En realidad, esta sustancia cuasicristalina también mostraba una simetría en *tres* dimensiones, y no sólo en el plano, dando globalmente una simetría icosaédrica prohibida (Shechtman y cols., 1984). (Robert Amman, en 1975, encontró "icosaedros" tridimensionales análogos a mis teselaciones planas quíntuples; véase Gardner, 1989.) Las aleaciones de Shechtman formaban sólo minúsculos cuasicristales microscópicos, de unos 10^{-3} milímetros de largo, pero posteriormente se descubrieron otras sustancias cuasicristalinas y, en particular, una aleación de aluminio-litio-cobre, para las que las unidades con simetría icosaédrica pueden

crecer hasta un tamaño de aproximadamente un milímetro y son perfectamente visibles a simple vista (véase fig. X.4).

Ahora bien, una característica notable de las estructuras teselantes cuasicristalinas que he estado describiendo es que su ensamblaje es necesariamente *no local.* Es decir, al ensamblar las estructuras es necesario, de cuando en cuando, examinar el estado de la estructura a muchos y muchos "átomos" de distancia del punto de ensamblaje si queremos estar seguros de no cometer graves errores cuando juntemos las piezas. (Esto es quizá parecido al aparente "andar a tientas inteligentemente" a que me referí en relación con la selección natural.) Este tipo de característica es un ingrediente de una controversia considerable que rodea actualmente a la cuestión de la estructura y crecimiento de los cuasicristales, y no sería prudente intentar sacar conclusiones definitivas hasta que algunos de los puntos más importantes queden resueltos De todas formas, podemos especular; y yo aventuraré mi propia opinión En primer lugar, creo que algunas de estas sustancias cuasicristalinas están fuertemente organizadas, y sus disposiciones atómicas tienen estructura muy próxima a las estructuras teselantes que he venido considerando. En segundo lugar, soy de la opinión (más tentativa) de que esto implica que un ensamblaje no puede conseguirse razonablemente mediante la adición local de átomos de uno en uno, de acuerdo con la imagen *clásica* del crecimiento cristalino, sino que en su lugar debe haber en su ensamblaje un ingrediente mecánico-cuántico esencialmente *no local.* [7]

El modo en que yo imagino que tiene lugar este crecimiento es que, en lugar de tener átomos que llegan individualmente y se acoplan a una línea de crecimiento en continuo movimiento (crecimiento cristalino clásico), debemos considerar la evolución de una superposición lineal cuántica de muchas diferentes configuraciones posibles de átomos que se acoplan (mediante el procedimiento cuántico **U**). De hecho, esto es lo que la mecánica cuántica nos dice que *debe* ocurrir (casi siempre). No se trata de que sólo suceda una cosa; muchas disposiciones atómicas posibles deben coexistir en una superposición lineal compleja. Unas cuantas de estas posibilidades superpuestas crecerán hasta constituir conglomerados más grandes y, en un cierto punto, la diferencia entre los campos gravitatorios de algunas de las posibilidades alcanzarán el nivel de un gravitón (o cualquiera que sea apropiado; véase capítulo VIII). En esta etapa, una de las configuraciones posibles —o, más exactamente, todavía una superposición, aunque una superposición algo reducida— se singulizará con la configuración "real" (procedimiento cuántico **R**). Este ensamblaje superpuesto, junto con reducciones a configuraciones más definidas, continuará a escala cada vez mayor hasta que se forma un cuasicristal de tamaño razonable.

Normalmente, cuando la naturaleza busca una configuración cristalina está buscando una configuración de *mínima energía* (suponiendo que la temperatura ambiente sea cero). Me imagino algo similar en el crecimiento de cuasicristales, con la diferencia de que este estado de mínima energía es mucho más difícil de encontrar, y la "mejor" configuración de átomos no *puede* descubrirse simplemente añadiendo átomos de uno en uno con la esperanza de poder considerar cada átomo por separado resolviendo simplemente su *propio* problema de minimización. En lugar de ello, tenemos que resolver un problema *global.* Debe haber un esfuerzo cooperativo de un gran número de átomos a la vez. Semejante cooperación, supongo, debe conseguirse mecánico-cuánticamente; y el modo de hacerlo es "ensayando"

[7] Incluso el crecimiento de ciertos cristales podría entrañar problemas similares, por ejemplo, en los casos en que la unidad celular básica comprende varios cientos de átomos, como en las llamadas fases de Frank-Casper. Debería mencionarse, por el contrario, que Onoda, Steinhardt y Socolar (1988) han propuesto un procedimiento de crecimiento teórico "casi local" (aunque todavía no local) para la simetría quíntuple de cuasicristales.

simultáneamente diferentes configuraciones de átomos combinadas en una superposición lineal (un poco quizá como la computadora cuántica considerada al final del capítulo IX). La selección de una solución apropiada (aunque probablemente no la mejor) al problema de la minimización debe conseguirse cuando se alcanza el criterio de un gravitón (u otro adecuado), lo que presumiblemente sólo ocurrirá cuando se den las condiciones físicas precisas.

POSIBLES REPERCUSIONES EN LA PLASTICIDAD CEREBRAL

Permítaseme llevar estas especulaciones un poco más lejos y preguntar si pudieran tener alguna importancia para la cuestión del funcionamiento del cerebro. Lo más probable, por lo que puedo ver, es que la tengan en el fenómeno de la plasticidad cerebral. Recordemos que el cerebro no es muy parecido a una computadora sino que se parece más a una computadora que está cambiando continuamente. Estos cambios pueden darse aparentemente por la activación o desactivación de sinapsis por medio del crecimiento o contracción de espinas dendríticas (véase capítulo IX, fig. IX. 15). Me atrevo a especular que este crecimiento o contracción podría estar gobernado por algo semejante al proceso que interviene en el crecimiento de los cuasicristales. En tal caso, no sólo se ensaya una, sino un inmenso número de posibles configuraciones, todas superpuestas de manera lineal y compleja. Mientras los efectos de estas configuraciones se mantengan por debajo del nivel de un gravitón (o cualquiera que sea) entonces ellas coexistirán (y casi invariablemente *deben* coexistir según las reglas de la mecánica cuántica **U**). Sin se mantienen por debajo de este nivel, pueden empezar a ejecutarse cálculos superpuestos simultáneos, de un modo que está en concordancia con los principios de una computadora cuántica. Sin embargo, parece improbable que estas superposiciones puedan mantenerse durante mucho tiempo, ya que las señales nerviosas producen campos eléctricos que perturbarían considerablemente el material circundante (aunque su vaina de mielina podría ayudar a aislarlas). Conjeturemos que tales superposiciones de cálculos pueden mantenerse realmente durante al menos el tiempo suficiente para calcular algo que sea de importancia antes de que alcance el nivel de un gravitón (o cualquiera que sea). El resultado acertado de tal cálculo sería el "objetivo" que toma el lugar del "objetivo" de la simple minimización de la energía en el crecimiento del cuasicristal. Por lo tanto la consecución de dicho similar al crecimiento con éxito del cuasicristal.

Hay obviamente mucha vaguedad e indecisión en estas especulaciones, pero creo que representan una analogía razonable. El crecimiento de un cristal o un cuasicristal está fuertemente influido por las concentraciones de los átomos y iones apropiados en su vecindad. Análogamente, se podrá concebir que el crecimiento o contracción de familias de espinas dendríticas podría perfectamente estar influido por las concentraciones de las diversas sustancias neurotransmisoras que estuvieran alrededor (tanto como podría estar afectado por las emociones). Cualesquiera que sean las configuraciones atómicas que finalmente se resuelvan (o "reduzcan") como la realidad del cuasicristal, ellas suponen la solución de un problema de minimización de la energía. De modo análogo, conjeturo que el pensamiento que sale a la superficie del cerebro es de nuevo la solución de cierto problema, aunque ahora no sólo es un problema de minimización de energía. Implicaría en general un objetivo de naturaleza mucho más compleja, que incluye deseos e intenciones en sí mismos relacionados con los aspectos y capacidades computacionales del cerebro. Se me figura que la acción del pensamiento consciente está muy ligada la resolución de configuraciones que previamente estaban en la superposición lineal. Todo esto está relacionado con la física desconocida que gobierna la línea divisoria entre

U y **R** y que, a mi modo de ver depende de una teoría de la gravitación cuántica —GQC— aún por descubrir.

¿Podría semejante acción física ser de naturaleza no algorítmica? Recordemos que el problema general de la teselación, como se describió en el capítulo IV, no tiene solución algorítmica. Se podría concebir que los problemas del ensamblaje de átomos podrían compartir esta propiedad no algorítmica. Si estos problemas pueden "resolverse" en principio por el tipo de medios que he estado insinuando, entonces existe alguna posibilidad para un ingrediente no algorítmico en el tipo de acción cerebral que tengo en mente. Para que esto sea así, sin embargo, necesitamos algo no algorítmico en la GQC. Evidentemente hay aquí mucha especulación. Pero *algo* de carácter no algorítmico me parece definitivamente necesario a la vista de los argumentos antes expuestos.

¿Con qué rapidez tienen lugar estos cambios en la conexión cerebral? La cuestión parece ser algo controvertida entre los neurofisiólogos pero, puesto que la memoria permanente puede asentarse en unas pocas fracciones de segundo, es posible que tales cambios en las conexiones puedan efectuarse en un tiempo de ese orden. Para la realización de mis propias ideas sería necesaria una velocidad así.

LOS RETARDOS TEMPORALES DE LA CONCIENCIA

A continuación quisiera describir dos experimentos (mostrados en Harth, 1982) que se han realizado en sujetos humanos y que al parecer tienen repercusiones notables en nuestras consideraciones sobre este punto. Tienen que ver con el tiempo que necesita la conciencia para actuar y ser activada. El primero de estos concierne al papel activo de la conciencia, y el segundo a su papel pasivo. Consideradas juntas, estas repercusiones son aún más sorprendentes.

El primero lo realizaron H. H. Kornhuber y sus colaboradores en Alemania en 1976. (Deecke, Grótzinger y Kornhuber, 1976.) Cierto número de personas se prestaron voluntariamente a que se registraran las señales eléctricas de un punto de su *cabeza* (mediante electroencefalogramas, esto es, EEG), y se les pedía que flexionaran varias veces, y repetidamente, el dedo índice de su mano derecha *a su propia elección.* Se partía del supuesto de que los registros de los EEG indicarían algo de la actividad mental que tenía lugar dentro del cráneo y se relaciona con la decisión consciente de flexionar el índice. Para obtener una señal significativa de las trazas del EEG es necesario promediar las trazas de varias series diferentes, y la señal resultante no es muy definida. Sin embargo, se encuentra algo curioso: hay un aumento gradual del potencial eléctrico registrado durante un segundo entero, o quizá incluso hasta un segundo y medio, *antes* de que el dedo sea flexionado. Esto parece indicar que el proceso de decisión consciente necesita un segundo para actuar. Esto puede contrastarse con el tiempo mucho más corto que lleva responder a una señal externa si el modo de respuesta ha sido establecido por adelantado. Por ejemplo, la flexión del dedo podría ser la respuesta al destello de una señal luminosa en lugar de darse "a voluntad". En tal caso es normal un tiempo de reacción de aproximadamente un quinto de segundo, que es unas cinco veces más rápido que la acción "voluntaria" que se ponía a prueba en los datos de Kornhuber (véase la fig. X.5)

En el segundo experimento, Benjamin Libet, de la Universidad de California, en colaboración con Bertram Feinstein del Instituto Neurológico del Monte Sión en San Francisco (Libet *et al,* 1979), ponían a prueba sujetos que tenían que sufrir intervención quirúrgica en el cerebro por

alguna razón que no tuviera relación con el experimento y consintieron en tener electrodos colocados en puntos de la corteza somatosensorial del cerebro.

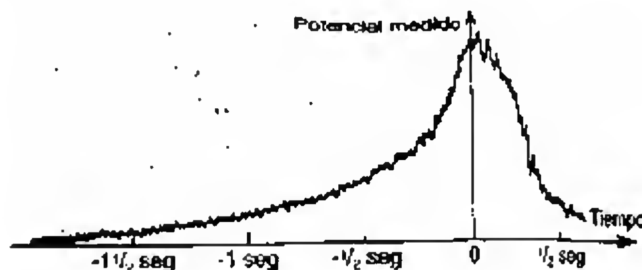


FIGURA X.5. *Experimento de Kornhuber. La decisión de flexionar el dedo parece hacerse en el tiempo O, pese a que la señal precursora (promediada sobre muchos ensayos) sugiere una "anticipación" de la intención de flexionar.*

El resultado del experimento de Libet fue que, cuando se aplicaba un estímulo en la piel de estos pacientes, transcurría aproximadamente medio segundo antes de que fueran conscientes de dicho estímulo, pese al hecho de que el propio cerebro había recibido la señal del estímulo en sólo una centésima de segundo y podría lograrse una respuesta del cerebro a dicho estímulo *(cfr. supra)* en aproximadamente una décima de segundo (fig. X.6). Además, a pesar del retardo de medio segundo antes de que el estímulo alcanzase la conciencia, existía la impresión subjetiva de los propios pacientes de que no había habido ningún retraso en su toma de conciencia del estímulo. (Algunos de los experimentos de Libet implicaban estimulación del tálamo, con resultados similares a los de la corteza somatosensorial.)

Recordemos que la corteza somatosensorial es la región del cerebro por la que entran las señales nerviosas. Por lo tanto, la estimulación eléctrica de un punto de ella, correspondiente a un punto particular en la piel, sería para el sujeto igual que si algo hubiera tocado realmente la piel en el punto correspondiente. Sin embargo, resulta que si la estimulación eléctrica es demasiado breve —durante menos de aproximadamente medio segundo— entonces el sujeto no llega a ser consciente de ninguna sensación. Esto debe contrastarse con una estimulación directa del punto de la misma piel, puesto que un contacto momentáneo en la piel sí puede dar lugar a sensación.

Supongamos ahora que se toca primero la piel y luego se estimula eléctricamente el punto de la corteza somatosensorial. ¿Qué siente el paciente? Si la estimulación eléctrica se inicia aproximadamente un cuarto de segundo después de tocar la piel, entonces el contacto en la piel no se siente en absoluto. Este es un efecto conocido como *enmascaramiento retroactivo.*

Por lo visto, la estimulación de la corteza sirve para impedir que la sensación normal de contacto en la piel sea sentida conscientemente.

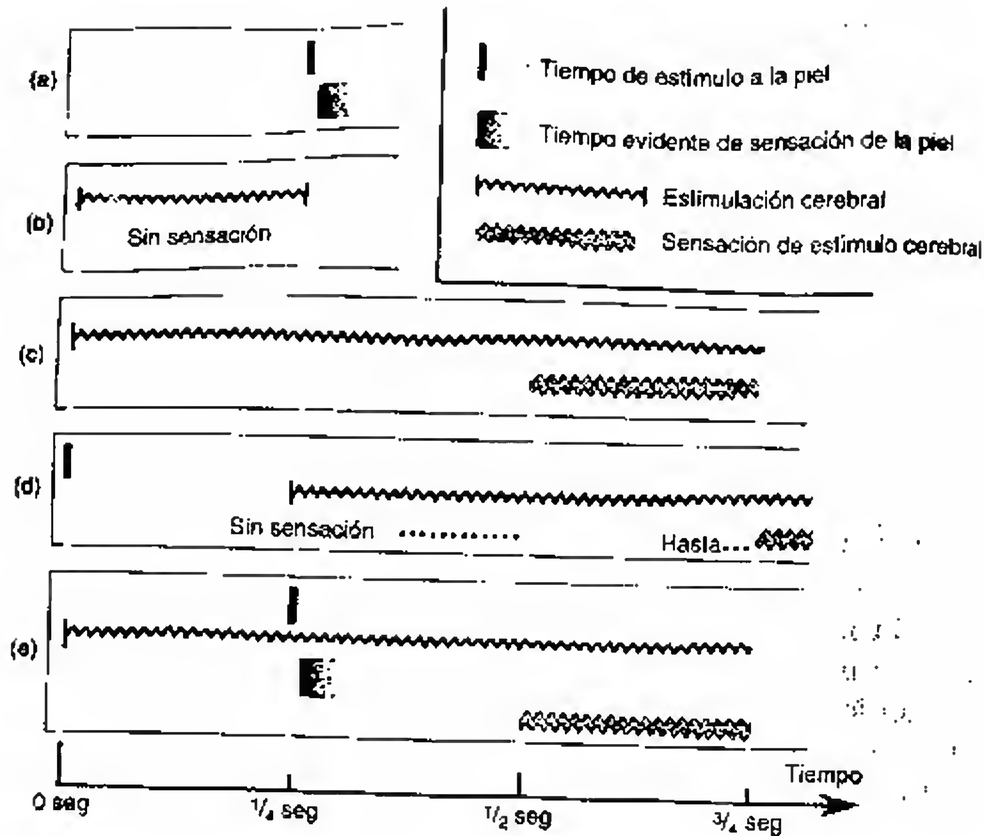


FIGURA X.6. *Experimento de Libet.* (a) *el estímulo en la piel "parece" percibirse aproximadamente en el momento real del estímulo,* (b) *Un estímulo cortical de menos de medio segundo no es percibido,* (c) *Un estímulo cortical de más de medio segundo se percibe a partir de medio segundo en adelante,* (d) *Tal estímulo cortical puede "enmascarar retroactivamente" un estímulo anterior en la piel, lo que indica que la conciencia del estímulo en la piel* no ha tenido lugar aún *en el momento del estímulo cortical,* (e) *Si un estímulo en la piel se aplica inmediatamente* después *del estímulo cortical, la percepción de la piel es "remitida hacia atrás" pero la percepción cortical no.*

La percepción consciente puede ser inhibida ("enmascarada") por un suceso posterior, siempre que dicho suceso ocurra antes de medio segundo. Esto nos dice que la percepción consciente de tal sensación ocurre como medio segundo después del suceso real que produce esa sensación.

Sin embargo no parece que seamos "conscientes" de un retardo temporal tan largo en nuestras percepciones. Una manera de hacer comprensible este curioso descubrimiento podría consistir en imaginar que el "tiempo" de todas nuestras "percepciones" está retrasado efectivamente alrededor de medio segundo respecto al "tiempo real" —como si el reloj interno de cada uno simplemente estuviese "equivocado" en medio segundo aproximadamente. Así, el momento en que se percibe que ocurre un suceso sería siempre medio segundo *posterior* a la presentación del mismo. Esto ofrecería una imagen consistente, si bien incómodamente retrasada, de las impresiones sensoriales.

Quizá algo de esta naturaleza queda corroborado en una segunda parte del experimento de Libet, en donde él iniciaba *primero* una estimulación eléctrica de la corteza, continuando esta estimulación durante un intervalo mucho mayor de medio segundo y tocando también la piel mientras seguía esta estimulación, aunque menos de medio segundo después de su inicio. Tanto la estimulación cortical como el contacto en la piel se percibían por separado, y resultaba claro para el sujeto cuál era cada uno. Sin embargo, cuando se le preguntaba qué estímulo ocurrió *primero,* el sujeto respondía que fue el contacto en la piel, pese al hecho de que la estimulación cortical se inició mucho antes. Por lo tanto, el sujeto parece remitir la percepción del contacto en la piel hacia *el pasado* en aproximadamente medio segundo (véase fig. X.6). Sin embargo, parece que esto no es simplemente un error "global" en el tiempo internamente percibido sino una readaptación de la percepción temporal de los sucesos, pues la estimulación cortical, suponiendo que se perciba realmente no más de medio segundo después de su inicio, *no* parece remitirse al pasado de esta forma.

A partir del primero de los experimentos anteriores parece deducirse que la acción consciente necesita algo así como un segundo o un segundo y medio antes de poder llevarse a cabo, mientras que, según el segundo experimento, la conciencia de un suceso externo no parece ocurrir hasta medio segundo después de que haya tenido lugar el suceso. Imaginemos lo que sucede cuando respondemos a algún acontecer externo imprevisto. Supongamos que la respuesta es algo que requiere un momento de contemplación consciente. Parecería, sobre la base de los experimento de Libet, que debiera transcurrir medio segundo antes de que la conciencia sea llamada a juego; y luego, como parecen implicar los datos de Kornhuber, se necesita bastante más de un segundo antes de que pueda tener efecto una respuesta "voluntaria". El proceso total, desde la entrada sensorial hasta la salida motriz, parecería necesitar un tiempo del orden de dos segundos. La consecuencia aparente de estos dos experimentos considerados juntos es que la conciencia no puede siquiera entrar en juego *en absoluto* en respuesta a un suceso externo, si dicha respuesta ha de darse en menos de un par de segundos aproximadamente.

EL EXTRAÑO PAPEL DEL TIEMPO EN LA PERCEPCIÓN CONSCIENTE

¿Podemos tomar estos experimentos al pie de la letra? Si es así, parece no haber más remedio que aceptar la conclusión de que actuamos completamente como "autómatas" cuando llevamos a cabo cualquier acción que necesite menos de un segundo o dos en modificar una respuesta. Sin duda la conciencia es de acción lenta comparada con otros mecanismos del sistema nervioso. Yo mismo me he encontrado en ocasiones observando inútilmente cómo mi mano cierra la puerta del coche un instante después de haber reparado en alguna cosa que hubiera querido recoger en el interior del coche, y mi orden voluntaria para detener el movimiento de mi brazo actúa de forma desesperantemente lenta, demasiado lenta para detener el cierre de la puerta. Pero ¿se requiere un segundo o dos para esto? Me parece poco probable que tenga que mediar un tiempo tan largo. Por supuesto, mi percepción *consciente* del objeto dentro del coche junto con mi supuesta "libre voluntad" de la orden para detener mi mano *podría* haber ocurrido perfectamente después de ambos sucesos. Quizá la conciencia es, después de todo, un espectador que no experimenta más que una "acción repetida" de la escena. De modo similar, y frente a esto, sobre la base de los descubrimientos anteriores no habría tiempo para que la conciencia desempeñara el más mínimo papel cuando, por ejemplo, ejecutamos un golpe de tenis y mucho menos de ping-pong. Sin duda los expertos en estas tareas tendrán todos los elementos esenciales de sus

respuestas soberbiamente preprogramadas en el control cerebelar. Pese a ello, encuentro un poco difícil dar crédito a que la conciencia *no* desempeñe el más mínimo papel en la decisión de qué golpe asestar en cada ocasión. Sin duda hay mucho de anticipación de lo que el adversario podría hacer, y muchas respuestas preprogramadas deberían estar disponibles para cada posible acción del adversario, pero me parece que esto es ineficaz y encontraría difícil aceptar una *total* ausencia de intervención consciente en ese instante. Comentarios semejantes serían aún más pertinentes en relación con la conversación ordinaria. También aquí, aunque pudiéramos prever en parte lo que el otro va a decir, debe haber normalmente algo inesperado en los comentarios del interlocutor o la conversación sería totalmente innecesaria. Ciertamente no se necesita tanto como un par de segundos para responder a alguien en el modo normal de conversación. Tal vez haya razón para dudar que los experimentos de Kornhuber demuestren que la conciencia necesita "realmente" un segundo y medio para actuar. Aunque es cierto que el *promedio* de todas las trazas del EEG para la intención de flexionar el dedo daba una señal tan prematura, pudiera ser que sólo en *algunos* casos haya una intención tan anticipada de flexionar el dedo —donde a menudo esta intención consciente *no* se materializa realmente— mientras que en muchos otros casos la acción consciente ocurra mucho más próxima a la flexión del dedo. (En realidad, algunos hallazgos experimentales —*cfr.* Libet, 1987, 1989—llevaron a interpretaciones un tanto diferentes de las de Kornhuber. Sin embargo, las intrigantes implicaciones con relación al tiempo de la conciencia siguen con nosotros)

Aceptemos, por el momento, que ambas conclusiones experimentales son realmente válidas. Me gustaría aventurar una hipótesis en relación con esto. Es probable que estemos equivocados al aplicar las reglas físicas usuales para el *tiempo* cuando consideramos la conciencia. Existe, efectivamente, algo muy singular en el modo en que el tiempo interviene en nuestras percepciones conscientes en cualquier caso, y pienso que es posible que sea necesaria una concepción muy diferente cuando tratamos de colocar las percepciones conscientes en un marco convencional de ordenación temporal. La conciencia es, después de todo, el único fenómeno que conocemos según el cual el tiempo necesita "fluir". El modo de considerar el tiempo en la física moderna no es esencialmente diferente de la manera en que se considera el *espacio,*[*] y el "tiempo" de las descripciones físicas no "fluye" en absoluto; sólo tenemos un "espacio-tiempo" fijo de apariencia estática en el que están dispuestos los sucesos de nuestro universo. No obstante, según nuestras percepciones, el tiempo *fluye (cfr.* capítulo VII). Mi conjetura es que aquí también existe algo ilusorio, y el tiempo de nuestras percepciones no fluye "realmente" en la forma de avance lineal en que lo percibimos fluir (independientemente de lo que esto pueda significar.) El ordenamiento temporal que uno "parece" percibir es, afirmo, algo que imponemos a nuestras percepciones para poder darles sentido en relación con la progresión temporal uniforme hacia adelante de una realidad física externa.

Algunas personas detectarán muchas "incorrecciones filosóficas" en los comentarios anteriores, y sin duda tendrán razón en estas acusaciones. ¿Cómo podemos estar "errados" sobre lo que realmente percibimos? Ciertamente, nuestras percepciones reales son, *por definición,* las únicas cosas de las que somos realmente conscientes; por lo tanto, no podemos estar "errados" sobre ellas. De todas formas, pienso que es probable que *estemos* "errados" sobre nuestras

[*] La simetría entre tiempo y espacio sería aún más sorprendente para un espacio-tiempo bidimensional. Las ecuaciones de la física en un espacio-tiempo bidimensional serían esencialmente simétricas con respecto al intercambio del espacio con el tiempo, pese a que nadie consideraría que el espacio "fluye" en la física de dos dimensiones. Es difícil creer que lo que hace que el tiempo "fluya realmente", en nuestras experiencias del mundo físico que conocemos, sea simplemente la asimetría entre el número de dimensiones espaciales (3) y temporales (1) que resulta tener nuestro espacio-tiempo.

percepciones del progreso temporal (a pesar de mis inexactitudes en el uso del lenguaje ordinario para describir esto) y que hay testimonios en favor de tal creencia *(véase* Churchland, 1984).

Un ejemplo extremo es la capacidad que tenía Mozart para "captar de golpe" una composición musical entera "por larga que pudiera ser". Debemos suponer, por la descripción de Mozart, que este "golpe" contenía la esencia de la composición entera, pese a que el intervalo de tiempo real, en términos físicos ordinarios, de este acto consciente de percepción no fuera en modo alguno comparable con el tiempo que se necesitaría para ejecutar la composición. Podríamos imaginar que la percepción de Mozart habría tomado una forma completamente diferente, quizá distribuida espacialmente como una esencia visual o toda una partitura musical desplegada. Pero incluso una partitura musical necesitaría un tiempo considerable para ser leída, y dudo mucho que la percepción de Mozart de sus composiciones pudiera haber tomado inicialmente esta forma (o seguro que lo habría dicho). La escena visual parecería más próxima a sus descripciones, pero (como las más comunes de las imágenes matemáticas que me son personalmente más familiares) dudaría mucho que existiera algo semejante a una traducción directa de la música en términos visuales. Me parece mucho más probable que la mejor interpretación del "golpe" debe hacerse *musicalmente,* con las connotaciones característicamente temporales que tendría la audición (o ejecución) de una pieza musical. La música consiste en sonidos cuya ejecución toma un tiempo definido, el *tiempo* que en la descripción real de Mozart permite "...que mi imaginación me deje oírla." Oigamos la cuádruple fuga en la parte final del *Arte de la fuga* de J. S. Bach. Nadie a quien le guste la música de Bach puede evitar conmoverse cuando la música se detiene después de diez minutos de ejecución, precisamente tras la entrada del tercer tema. La composición en su totalidad parece seguir "ahí", pero ahora se nos ha desvanecido en un instante. Bach murió antes de poder completar la obra y su partitura simplemente se detiene en este punto, sin ninguna indicación escrita de cómo pensaba continuarla. Pero empieza con tal seguridad y completo dominio que resulta imposible imaginar que no tuviese la esencia de la composición entera en su cabeza en ese instante. ¿Tendría necesidad de tocarla en su totalidad para sí mismo en su mente, al ritmo de ejecución normal, ensayando una y otra vez, a medida que se le ocurrían las diferentes mejoras? No puedo imaginar que lo hiciera en esta forma. Al igual que Mozart, él debió haber sido capaz de concebir la obra en su totalidad, con la intrincada complejidad y sentido artístico que exige la escritura fugada, evocada toda a la vez. Pero la cualidad temporal de la música es uno de sus ingredientes esenciales. ¿Cómo puede la música seguir siendo música si no se está ejecutando en "tiempo real"?

La concepción de una novela o una historia podría presentar un problema comparable, aunque aparentemente menos enigmático. Para la comprensión de la vida entera de un individuo necesitaríamos contemplar diversos sucesos cuya adecuada apreciación parecería requerir su actualización mental en "tiempo real". Pero no parece que esto sea necesario. Incluso las impresiones de los recuerdos de nuestras propias experiencias extendidas en el tiempo parecen estar de algún modo tan "comprimidas" que virtualmente podemos "revivirlas" en un instante de rememoración.

Hay quizá una estrecha semejanza entre la composición musical y el pensamiento matemático. La gente podría suponer que una demostración matemática se concibe como una progresión lógica, en donde cada paso se sigue de los precedentes. Pero es apenas probable que la concepción de un nuevo razonamiento proceda de esta forma. Existe una globalidad y un contenido conceptual aparentemente vago que son necesarios en la construcción de un argumento matemático; y esto puede guardar poca relación con el tiempo que parecería

necesitarse para apreciar completamente una demostración presentada seriamente. Supongamos, entonces, que aceptamos que el ritmo y la progresión temporal de la conciencia no concuerda con el de la realidad física externa. ¿No estamos en peligro de llegar a una paradoja? Supongamos que haya incluso algo vagamente teleológico en los efectos de la conciencia, de modo que una impresión futura pueda afectar a una acción pasada. ¿Nos llevaría *esto* a una contradicción, como las consecuencias paradójicas de las señales ultralumínicas que consideramos — y precisamente descartamos — hacia el final del capítulo V? *No* hay necesariamente una paradoja — por la propia naturaleza de lo que, según sostengo, consigue la conciencia. Recuérdese mi propuesta de que la conciencia es, en esencia, la "visión" de una verdad necesaria; y que puede representar un tipo de contacto real con el mundo platónico de los conceptos matemáticos ideales. Recuérdese que el mismo mundo platónico es intemporal. La percepción de la verdad platónica no supone información — en el sentido técnico de "información" que puede transmitirse mediante un mensaje — y no habría contradicción real envuelta si tal percepción consciente se propagara incluso hacia el pasado.

Pero aun si aceptamos que la propia conciencia tiene esa curiosa relación con el tiempo — y que representa, en cierto sentido, un contacto entre el mundo físico externo y algo intemporal — ¿cómo puede esto ajustarse a una acción físicamente determinada u ordenada temporalmente del cerebro material? Una vez más, parece que nos quedamos con un simple papel de "espectador" para la conciencia si no podemos reflejar la progresión normal de las leyes físicas. Pero mi argumentación va en favor de algún tipo de papel activo para la conciencia, y en realidad para un papel poderoso, con una fuerte ventaja selectiva. La respuesta a este dilema, creo yo, reside en la extraña forma en que deben actuar la GQC en su resolución del conflicto entre los dos procesos mecánico-cuánticos **U** y **R** .

Recuérdense los problemas con el tiempo que encuentra el proceso R cuando tratamos de hacerlo compatible con la relatividad especial (capítulos VI, VIII). El proceso no parece tener sentido en absoluto cuando se escribe en función del espacio-tiempo. Consideremos el estado cuántico de un par de partículas. Tal estado será normalmente un estado *correlacionado* (es decir, no de la simple forma $|\psi\rangle|\chi\rangle$—donde cada uno de los $|\psi\rangle$ y $|\chi\rangle$ describe sólo una de las partículas— sino una suma, como $|\psi\rangle|\chi\rangle + |\alpha\rangle|\beta\rangle + ... + |\rho\rangle|\sigma\rangle$. En tal caso, la observación en una de las partículas afectará a la otra de una forma no local que no puede describirse en función del espacio-tiempo ordinario de forma compatible con la relatividad especial (EPR; el efecto Einstein-Podolski-Rosen). Tales efectos no locales estarían implícitos en mi analogía entre el "cuasicristal" y el crecimiento y contracción de las espinas dendríticas. Aquí interpreto "observación" en el sentido de una amplificación de la acción de cada partícula observada hasta que se satisfaga algo parecido al criterio de un gravitón de GQC. En términos más "convencionales", una "observación" es una cosa mucho más oscura y es difícil ver cómo podría empezarse a desarrollar una descripción teórico-cuántica de la acción cerebral cuando perfectamente podríamos tener que considerar que el cerebro "se observa a sí mismo" continuamente.

Mi propia idea es que la GQC, por el contrario, proporciona una teoría física *objetiva* de la reducción del vector de estado (**R**) que *no* tendría que depender de ninguna idea de conciencia. Todavía no tenemos una teoría semejante, pero al menos su descubrimiento no será obstaculizado por los profundos problemas para decidir qué "es" realmente la conciencia.

Imagino que una vez que se haya descubierto finalmente la GQC podrá elucidarse en sus términos el fenómeno de la conciencia. En realidad creo que las propiedades exigidas a la GQC, cuando llegue la teoría, estarán incluso más alejadas de una descripción espacio-temporal convencional de lo que están los enigmáticos fenómenos EPR para las dos partículas citadas antes. Si, como afirmo, el fenómeno de la conciencia depende de esta deseada GQC, entonces la propia conciencia se adaptará sólo de forma bastante incómoda a nuestras actuales descripciones espacio-temporales.

CONCLUSIÓN: LA VISIÓN DE UN NIÑO

En este libro he presentado muchos argumentos para mostrar lo insostenible del punto de vista —aparentemente prevaleciente en la corriente filosófica actual— de que nuestro pensamiento es básicamente lo mismo que la acción de una computadora muy complicada. Se ha adoptado aquí la terminología de Searle "IA fuerte" para la suposición explícita de que la simple operación de un algoritmo puede provocar *conocimiento consciente.* Otros términos tales como "funcionalismo" son utilizados a veces de una manera algo menos definida.

Algunos lectores pueden haber considerado desde el principio que el "defensor de la IA fuerte" era más que nada un enemigo ficticio. ¿No es "obvio" que la simple computación no puede provocar placer o dolor; que no puede percibir la poesía, o la belleza del cielo al atardecer, o la magia de los sonidos; que no puede tener esperanza o amar o desesperar; que no puede tener un objetivo genuino autónomo? Pero la ciencia parece habernos llevado a aceptar que todos somos simplemente pequeñas partes de un mundo gobernado en todo detalle (incluso si finalmente resultara ser quizá de manera probabilista) por leyes matemáticas muy precisas. Nuestros propio cerebro, que parece controlar todas nuestras acciones, está gobernado también por estas mismas leyes precisas. Ha surgido la imagen de que toda esa actividad física es, en efecto, nada más que la activación de algún enorme cómputo (quizá probabilista), y por ello nuestro cerebro y nuestra mente tienen que ser comprendidos solamente en términos de semejantes cómputos. Quizá cuando los cómputos se vuelvan extraordinariamente complicados puedan empezar a adoptar las cualidades más poéticas o subjetivas que asociamos al término "mente". No obstante, es difícil evitar un sentimiento incómodo de que siempre se echará algo de menos en esta imagen.

En mis propios argumentos he tratado de apoyar esta idea de que debe haber algo esencial que está ausente de cualquier imagen puramente computacional. Pero mantengo también la esperanza de que es por medio de la ciencia y las matemáticas como deben salir a la luz algunos avances profundos en la comprensión de la mente. Existe aquí un aparente dilema, pero he tratado de mostrar que *existe* una auténtica salida. Computabilidad no es en modo alguno la misma cosa que ser matemáticamente exacto. Existe en el mundo matemático platónico y exacto tanto misterio y belleza como pudiéramos desear, y la mayor parte de este misterio reside en conceptos que están fuera de esa parte relativamente limitada en la que residen los algoritmos y la computación.

La conciencia me parece un fenómeno de tal importancia que no puedo creer que sea algo producido "accidentalmente" por un cómputo complicado: es el fenómeno en el que se hace conocida la misma existencia del Universo. Podemos argumentar que un universo gobernado por leyes que no permiten la conciencia no es universo en absoluto. Diría incluso que todas las descripciones matemáticas del Universo que se han dado hasta ahora deben incumplir este

criterio. Es sólo el fenómeno de la conciencia el que puede conjurar un presunto universo "teórico" a la existencia real.

Algunos de los argumentos que he dado en estos capítulos pueden parecer tortuosos y complicados. Algunos son francamente especulativos, aunque yo creo que no existe escapatoria real entre unos u otros. Pero por encima de estos tecnicismos está el sentimiento de que es realmente "obvio" que la mente *consciente* no puede trabajar como una computadora, ni siquiera pese a que mucho de lo que realmente interviene en la actividad mental podría hacerlo.

Este es el tipo de obviedad que puede ver un niño, aunque ese niño pueda, en su vida posterior, verse llevado a creer que los problemas obvios son "no problemas" cuya no existencia se argumenta mediante cuidadosos razonamientos y definiciones inteligentemente escogidas. A veces los niños ven más claramente las cosas que se van oscureciendo con el avance de la vida. Cuando las preocupaciones de la "vida real" han comenzado a cargarse sobre nuestros hombros solemos olvidar las maravillosas sensaciones de cuando éramos niños. Los niños no tienen miedo de hacer preguntas básicas que nosotros, como adultos, sentiríamos vergüenza de plantear. ¿Qué sucede a cada uno de nuestros flujos de conciencia después de morir?, ¿dónde estaba antes de que naciera cada uno?, ¿podríamos convertirnos en otro, o haber sido algún otro? ¿por qué percibimos?, ¿por qué estamos aquí? ¿por qué hay un universo en el que podamos estar? Estos son enigmas que tienden a llegar con el despertar de la conciencia en cualquiera de nosotros —y, sin duda, con el despertar de la verdadera autoconciencia, dentro de cualquier criatura u otra entidad en que primero llegue.

Recuerdo que muchos de estos enigmas me inquietaban cuando era niño. Quizá mi propia conciencia podía intercambiarse repentinamente con la de algún otro. Cómo saber si no pudo haberme sucedido tal cosa antes, suponiendo que cada persona porta sólo sus propios recuerdos importantes? ¿Cómo podría explicar semejante experiencia de "intercambio" con algún otro? ¿Realmente significa algo? Quizá estoy viviendo simplemente las mismas experiencias de diez minutos una y otra vez, cada vez con las mismas percepciones exactamente. Quizá sólo "existe" para mí el instante presente. Quizá el "yo" de mañana, o el de ayer, es realmente una persona muy diferente con una conciencia independiente. Quizá estoy viviendo realmente hacia el pasado, con mi flujo de conciencia marchando hacia atrás de modo que mi memoria me dice más de lo que *va a suceder* que de lo *sucedido,* de modo que la experiencia desagradable en la escuela es realmente algo que me aguarda y, desgraciadamente, encontraré en muy breve plazo.

¿"Significa" algo la distinción entre eso y la progresión temporal experimentada, de modo que una es "falsa" y la otra "correcta"? Para que las respuestas a semejantes preguntas sean resueltas en principio, sería necesaria una teoría de la conciencia. Pero ¿cómo podríamos siquiera *empezar* a explicar la sustancia de tales problemas a una entidad que no sea en sí misma consciente...?

EPÍLOGO

"...LO QUE SE SIENTE SER...? Oh, ...una pregunta muy interesante, muchacho... eh... Yo también quisiera saber la respuesta", dijo el diseñador en jefe. "Veamos lo que nuestro amigo tiene que decir a eso... Esto es extraño... eh... Ultronic dice que él no ve..., ni siquiera puede comprender lo que planteas!" Los murmullos de risas por la sala estallaron en grandes carcajadas.

Adam se sintió molesto en extremo. Podían haber hecho cualquier otra cosa, pero no reírse.

REFERENCIAS

Aharonov, Y. y D. Z. Albert (1981), "Can we make sense out of the measurement process in relativistic quantum mechanics?" *Phys Rev* D24 pp. 359-370.

___, P. Bergmann y J. L. Lebowit (1964), "Time symmetry in the quantum process of measurement", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983 originalmente aparecido en *Phys. Rev.* 134B, pp. 1410-1416.

Ashtekar, A., A. P. Balachandran y Sang Jo (1989), "The CP problem in quantum gravity", *Int. J. Mod. Phys.,* A6, pp. 1493-1514.

Aspect, A. y P. Grangier (1986), "Experiments on Einstein-Podolsky-Rosen type correlations with pairs of visible photons", en R. Penrose y C. J. Isham (comps.), *Quantum concepts in space and time* Oxford University Press.

Atkins, P. W. (1987), *Why mathematics works,* Oxford University Extension Lecture in series: Philosophy and the New Physics (13 de marzo).

Barbour, J. B. (1989), *Absolute or relative motion?,* volumen I: *The discovery of dynamics,* Cambridge University Press, Cambridge.

Barrow, J. D. (1988), *The world within the world,* Oxford University Press.

____y F. J. Tipler (1986), *The anthropic cosmological principle,,* Oxford University Press.

Baylor, D. A., T. D. Lamb y K.-W. Yau (1979), "Responses of retinal rods to single photons", *J.Physiol,* 288, pp. 613-634.

Bekenstein, J. (1972), "Black holes and entropy", *Phys. Rev.,* D7, pp. 2333-2346.

Belinfante, F. J. (1975), *Measurement and time reversal in objective quantum theory,* Pergamon Press, Nueva York.

Belinskii, V. A., I. M. Khalatnikov y E. M. Lifshitz (1970), "Oscillatory approach to a singular point in the relativistic cosmology", *Adv. Phys.* 19, pp. 525-573.

Bell, J. S. (1987), *Speakable and unspeakable in quantum mechanics,* Cambridge University Press.

Benacerraf, P. (1967), "God, the Devil and Gödel", *The Monist,* 51, pp. 9-32.

Blakemore, C. y S. Greenfield (comp.) (1987), *Midwaves: thoughts on intelligence. Identity and consciousness,* Basil Blackwell, Oxford.

Blum L., M. Shub y S. Smale (1989), "On a theory of computation and complexity over the real numbers: NP completeness, recursive functions and universal machines", *Bull. Amer. Math. Soc.* (New series) 21, pp. 1-46.

Bohm, *D.* (1951), "The Paradox of Einstein, Rosen and Podolsky", en J.A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido en *Quantum theory,* D. Bohm, cap. 22, sec. 15-19. Prentice-Hall, Englewood-Clifts.

____(1952), "A suggested interpretation of the quantum theory in terms of ¨hidden¨ variables", I y II, en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido en *Phys. Rev.,* 85, pp. 166-193.

Bondi, H. (1960), "Gravitational waves in general relativity", *Nature,* Londres 186, p. 535.

Bowie, G. L. (1982), "Lucas´ number is finally up", *J. of Philosophical Logic,* 11, pp. 279-285.

Cartan, É. (1923), "Sur les varietés, á connexion affine et la théorie de la relativité généralisée", *Ann. Sci. EC. Norm. Sup.* 40, pp. 352-412.

Chandrasekhar, S. (1987), *Truth and beauty: aesthetics and motivations in science,* University of Chicago Press.

Church, A. (1941), "The calculi of lambda-conversion", *Annals of Mathematicals Studies,* num. 6, Princeton University Press.

Churchland, P. M. (1984), *Matter and consciousness,* Bradford Books, MIT Press, Cambridge, Mass.

Clauser, J. F., A. H. Home, A. Shimony y R. A. Holt (1969), "Proposed experiment to test local hidden-variable theories", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido en *Phys. Rev. Lett.* 23, pp. 880-884.

Close, F. (1983), *The cosmic onion: quarks and the nature of the universe,* Heinemann, Londres. (Hay traducción española: *La cebolla cósmica,* Editorial Crítica, Barcelona, 1988.)

Cohen, P. C. (1966), *Set theory and the continuum hypothesis,* Benjamin, Menlo Park, California.

Cutland, N. J. (1980), *Computability: an introduction to recursive function theory,* Cambridge University Press.

Davies, P. C. W. (1974), *The physics of time asymmetry,* Surrey University Press.

____y J. Brown (1988), *Superstrings ¿a theory of everything?,* Cambridge

University Press. (Hay traducción española: *Supercuerdas de todo?* Alianza Editorial, Madrid, 1990.)

Davies, R. D., A. N. Lasenby, R. A. Watson, E. J. Daintree J Ho W J. Beckman, J. Sanchez-Almeida y R. Rebolo (1987), "Sensitive measurement of fluctuations in the cosmic microwave background" *Nature,* 326, pp. 462-465.

Davis, M. (1988), "Mathematical logic and the origin of modern computers", en R. Herken (comp.), *The universal Turing machine: a half-century survey,* Kammerer & Unverzagt, Hamburgo.

Dawkins, R. (1986), *The blind watchmaker,* Longman, Londres. (Hay traducción española: *El relojero ciego,* Editorial Labor, Barcelona, 1989)

De Broglie, L. (1956), *Tentative d'interpretation caúsale et nonlinéaire de la méchanique ondulatoire,* Gauthier-Villars, París.

Deeke, L., B. Grótzinger y H. H. Kornhuber (1976), "Voluntary finger movements in man: cerebral potentials and theory", *Biol. Cybernetics,* 23, p. 99.

Delbrück, M. (1986), *Mind from matter?* Blackwell Scientific Publishing, Oxford.

Dennet, D. C. (1978), "Brainstorms", *Philosophical Essays on Mind and Psychology,* Harvester Press, Hassocks, Sussex.

Deutsch, D. (1985), "Quantum theory, the Church-Turing principle and the universal quantum computer", *Proc. Roy. Soc.* (Lond.), A400, pp. 97-117.

Devlin, K. (1988), *Mathematics: the new golden age,* Penguin Books, Londres.

De Witt, B. S, y R. D. Graham (comp.) (1973), *The many worlds interpretation of quantum mechanics,* Princeton University Press.

Dirac, P. A. M. (1928), "The quantum theory of the electron", *Proc. Roy. Soc.* (Lond.), A117, pp. 610-624; *ditto,* parte II, *ibid.,* A118, p. 361.

____(1938), "Classical theory of radiating electrons", *Proc. Roy. Soc.* (Lond.), A167, p. 148.

____(1939), "The relations between mathematics and physics", *Proc. Roy. Soc. Edinburgh,* 59, p. 122.

____(1947), *The principles of quantum mechanics* (3a. ed.), Oxford University Press. (Hay traducción española: *Principios de mecánica cuántica,* Editorial Ariel, Barcelona, 1958.)

____(1982), "Pretty mathematics", *Int. J. Theor. Phys.,* 21 pp. 603-605.

Drake, S. (trad.) (1953) *Galileo Galilei: dialogue concerning the two chief world systems — Ptolemaic and Copernican,* University of California, Berkeley, 1973. (Hay traducción española de los diálogos de Galileo: *Diálogo sobre los sistemas máximos,* traducción, prólogo y notas de José Manuel Revuelta, Aguilar, Buenos Aires, 1975.)

____(1957), *Discoveries and opinions of Galileo,* Doubleday, Nueva York.

Eccles J C. (1973), *The understanding of the brain,* McGraw-Hill, Nueva York

Einstein, A., B. Podolsky y N. Rossen (1935), "Can quantum-mechanical description of physical reality be considered complete?", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum mechanics and measurement,* Princeton University Press, 1983; originalmente aparecido en *Phys. Rev.* 47, pp. 777-780.

Everett, H . (1957), "'Relative State' Formulation of quantum mechanics", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido *en Rev. of Mod. Phys., 29,* pp. 454-462.

Feferman, S. (1988), "Turing in the Land of O(z)", en R. Herken (comp.), *The universal Turing machine: a half-century survey,* Kammerer & Unverzagt, Hamburgo.

Feynman, R. P. (1985), *QED: The strange theory of light and matter,* Princeton University Press. (Hay traducción española: *Electrodinánica cuántica. La extraña teoría de la luz y la materia,* Alianza Editorial, 1988.)

____, R. B. Leighton y M. Sands (1965), *The Feynman Lectures,* Addison Wesley. (Hay edición bilingüe inglés-español en Fondo Educativo Interamericano, Panamá, 1972.)

Fodor, J. A. (1983), *The modularity of mind,* MIT Press, Cambridge, Mass.

Fredkin, E. y T. Toffoli (1982), "Conservative logic", *Int. J. Theor. Phys.,* 21, pp. 219-253.

Freedman, S. J. y J. F. Clauser (1972), "Experimental test of local hidden-variable theories", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido en *Phys. Rev. Lett.,* 28, pp. 938-941.

Galilei, G. (1638), *Dialogues concerning two new sciences,* Macmillan, 1914, Dover Inc. (Hay traducción española: *Consideraciones y demostraciones matemáticas sobre dos nuevas ciencias,* ed. por C. Solís y J. Sádaba, Editora Nacional, Madrid, 1976.)

Gandy, R. (1988), "The Confluence of Ideas in 1936", en R. Herken (comp.), *The universal Turing machine: a half-century survey,* Kammerer & Unverzagt, Hamburgo.

Gardner, M. (1958), *Logic machines and diagrams,* University of Chicago Press. (Hay traducción española: *Máquinas y diagramas lógicos,* Alianza Editorial, Madrid, 1985.)

____(1983), *The whys of a philosophical scrivener,* William Morrow and Co., Inc., Nueva York. (Hay traducción española: *Los porqué de un escriba filosofo,* Editorial Tusquets, Barcelona, 1989.)

____(1989), *Penrose tiles to trapdoor ciphers,* W. H Freeman and Company, Nueva York.

Gayle, F. W. (1987), "Free-surface solidification habit and point symmetry of a faceted icosahedral Al-Li-Cu phase", *J. Mater* Res 2 pp. 1-4.

Gazzaniga, M. S. (1970), *The bisected brain,* Appleton-Century-Crofts, Nueva York.

____, J. E. LeDoux y D. H. Wilson (1977), "Language, praxis, and the right hemisphere: clues to some mechanisms of consciousness" *Neurology, 27,* pp. 1144-1147

Geroch, R. y J. B. Hartle (1986), "Computability and physical theories", *Found. Phys.,* 16, p. 533.

Ghirardi, G. C., A. Rimini y T. Weber (1980), "A general argument against superluminal transmission through the quantum mechanical measurement process", *Lett. Nuovo Chim.,* 27, pp. 293-298.

_____, (1986), "Unified dynamics for microscopic and macroscopic systems", *Phys. Rev.,* D34, p. 470,

Gödel, K. (1931), "Über formal unentscheidbare Sátze der Principia Mathematica und verwandter Systeme I", *Monatshefte fur Mathematik una Physik,* 38, pp. 173-198. (Hay traducción española en *Cuadernos Teorema,* Universidad de Valencia, 1981; y también en K. Gödel, *Obras completas,* Alianza Editorial, Madrid, 1987.)

Good, I. J. (1969), "Gödel's theorem is a red herring", *Brit. J. Philos Sci.,* 18, 359-373.

Gregory, R. L. (1981), *Mind in science: A history of explanations in psychology and physics,* Weidenfeld and Nicholson Ltd.

Grey Walter, W. (1953), *The living brain,* Gerald Duckworth and Co. Ltd.

Grünbaum, B. y G. C. Shephard (1981), "Some problems in plane tilings", en D. A. Klarner (comp.), *The mathematical Gardner,* Prindle, Weber and Schmidt, Boston.

_____, (1987), *Tilings and patterns,* W. H. Freeman.

Hadamard, J. (1945), *The psychology of invention in the mathematical field,* Princeton University Press. (Reeditado en 1954 en Dover Publishing Ltd., el informe completo de Poincaré

"La invención matemática" puede encontrarse en *Ciencia y método,* Espasa Calpe, Madrid, 1968 [N. del T.].)

Hanf, W. (1974), "Nonrecursive tilings of the plane, I",/. *Symbolic Logic,* 39, pp. 283-285.

Harth E. (1982), *Windows of the Mind,* Harvester Press, Hassoks, Sussex.

Hartle J B y S.W Hawking (1983), "Wave function of the universe", Phys rev., D31 p.1777

Hawking S.W (1975), "Particle creation by black holes", *Commun. Math Phys,* 43, PP. 199-220.

___ (1987), "Quantum cosmology", en S. W. Hawking y W. Israel (comps.), *300 years of gravitation,* Cambridge University Press. .

____(1988), *A brief history of time,* Bantam Press, Londres. (Hay traducción española: *Historia del tiempo,* Editorial Crítica, Barcelona, 1988; y Alianza Editorial, Madrid, 1990.)

____y R. Penrose (1970), "The singularities of gravitational collapse and cosmology", *Proc. Roy. Soc.* (Lond.), A134, pp. 529-548.

Hebb, D. O. (1954), "The problem of consciousness and introspection", en J. F. Delafresnaye (comp.), *Brain mechanisms and consciousness,*

Blackwell, Oxford. Hecht, S., S. Shlaer y M. H. Pirenne (1941), "Energy, quanta and vision", *J. of Gen. Physiol, 25,* pp. 891-840.

Herken, R. (comp.) (1988), *The universal Turing machine: a half-century survey,* Kammerer & Unverzagt, Hamburgo.

Hiley, B. J. y F. D. Peat, (comps.) (1987), *Quantum implications. Essays in honour of David Bohm,* Routledge and Kegan Paul, Londres y Nueva York.

Hodges, A. P. (1983), *Alan Turing: the enigma,* Burnett Books and Hutchinson, Londres; Simon and Schuster, Nueva York.

Hofstadter, D. R. (1979), *Gödel. Escher, Bach: an eternal golden braid,* Harvester Press, Hassoks, Sussex. (Hay traducción española: *Gödel, Escher, Bach,* Editorial Tusquets, Barcelona, 1989.)

____(1981), "A conversation with Einstein's brain", en D. R. Hofstadter y D. C. Dennett (comps.), *The mind's I,* Basic Books, Inc.; Penguin Books, Ltd; Harmondsworth, Middx.

____y Dennett, D. C. (comps.) (1981), *The mind's I,* Basic Books, Inc.; Penguin Books, Ltd; Harmondsworth, Middx.

Hubel, D. H. (1988), *Eye, brain and vision,* Scientific American Library Series, num. 22.

Huggett, S. A. y K. P. Tod (1985), *An introduction to twistor theroy,* Londres Math. Soc. student texts, Cambridge University Press.

Jaynes, J., (1980), *The origin of consciousness in the breakdown of the bicameral mind,* Penguin Books Ltd.; Harmondsworth, Middx.

Kandel, E. R. (1976), *The cellular basis of behaviour,* Freeman, San Francisco.

Károlyházy, F. (1974), "Gravitation and quantum mechanics of macroscopic bodies", *Magyar Fizikai Folyóirat,* 12, p. 24.

____ A. Frenkel y B. Lukács (1986), "On the possible role of gravity on the reduction of the wave function", en R. Penrose y C.J. Isham (comps) *Quantum concepts in space and time,* Oxford University Press

Keene, R. (1988), "Chess: Henceforward", *The Spectator,* 261, num 8371 p52

Knuth, D. M. (1981), *The art of computer programming,* vol. 2 (2a edición) Addison-Wesley, Reading, MA.

Komar, A. B. (1964), "Undecidability of macroscopically distinguishable states in quantum field theory", *Phys. Rev.,* 133b, pp. 542-544

____(1969), "Qualitative features of quantized gravitation" *Int J Theor. Phys.,* 2, pp. 157-160.

Kuznetsov, B. G. (1977), *Einstein: Leben, Tod. Unsterblichkeit* (traducido al alemán por H. Fuchs), Birkhauser, Basilea.

Le Doux, J. E. (1985), "Brain, mind and language", en D. A. Oakley (comp.), *Brain and mind,* Methuen, Londres y Nueva York.

Levy, D. W. L. (1984), *Chess computer handbook,* Batsford.

Lewis, D. (1969), "Lucas against mechanism", *Philosophy,* 44, pp. 231-233.

____(1989), "Lucas against mechanism II", *Can. J. ofPhilos., 9,* p. 373-376.

Libet, B., E. W. Jr. Wright, B. Feinstein y D. K. Pearl (1979), "Subjective referral of the timing for a conscious sensory experience", *Brain,* 102, pp. 193-224.

____(1987), "Consciousness: Conscious subjective experience", en *Enciclopaedia of neuroscience,* vol. 1, ed. por G. Adelman, Birkhauser, pp. 271-275.

____(1989), "Conscious subjective experience vs. unconscious mental functions: A theory of the cerebral process involved", en R. M. J. Cotterill (comp.), *Models of brain function,* Cambridge University Press, Cambridge, pp. 35-43.

Lorenz, K. (1972), citado en H. Wendt, *From ape to Adam,* Bobbs Merrill, Indianapolis.

Lucas, J. R. (1961), "Minds, machines and Gödel", *Philosophy,* 36, pp. 120-124; reimpreso en Alan Ross Anderson (1964), *Minds and machines,* Englewood Cliffs.

MacKay, D. (1987), "Divided brains-divided minds?", en C. Blakemore y S. Greenfield (comps.), *Mindwaves,* Basil Blackwell, Oxford.

Majorana, E. (1932), "Atomi orientati in campo magnético variabile", *Nuovo Cimento,* 9, pp. 43-50.

Mandelbrot, B. B. (1986), "Fractals and the rebirth of iteration theory", en H.-O. Peitgen y P. H. Richter, *The beauty of fractals: images of complex dynamical systems,* Springer-Verlag, Berlin, pp. 151-160.

Maxwell J. C. (1865), "A dynamical theory of the electromagnetic field", *Philos. Trans. Roy. Soc.* (Lond.), 155, pp. 459-512.

Mermin, D. (1985), "Is the moon there when nobody looks? Reality and the quantum theory", *Physics Today,* 38, num. 4, pp. 38-47.

Michie, D. (1988), "The fifth generation's unbridged gap", en R. Herken (comp.), *The universal Turing machine: a half-century survey* Kammerer & Unverzagt, Hamburgo.

Minsky, M. L. (1968), "Matter, mind, and models", en M. L. Minsky (comp.), *Semantic information processing,* M.I.T. Press, Cambridge, Mass.

Misner, C. W. (1969), "Mixmaster universe", *Phys. Rev. Lett.,* 22, pp. 1071-1074.

Moruzzi, G. y H. W. Magoun (1949), "Brainstem reticular formation and activation of the EEC", *Electroencephalography and Clinical Neuro-physiology,* 1, pp. 455-473.

Mott, N. F. (1929), "The wave mechanics of the c-tracks", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido en *Proc. Roy. Soc.* (Londres), A126, pp. 79-84.

____y Massey, H. S. W. (1965), "Magnetic moment of the electron", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983; originalmente aparecido en N. F. Mott y H. S. W. Massey, *The theory of atomic collisions,* Clarendon Press, Oxford, 1965.

Myers, D. (1974), "Nonrecursive tilings of the plane, II", *J. Symbolic Logic,* 39, pp. 286-94.

Myers, R. E. y R. W. Sperry (1953), "Interocular transfer of a visual form discrimination habit in cats after the section of the optic chiasm and corpus callosum", *Anatomical Record,* 175, pp. 351-352.

Nagel, E. y J. R. Newmann (1958), *Gödel's proff,* Routledge and Keegan Paul Ltd. (Hay traducción española: *El teorema de Gödel,* editorial Tecnos, Madrid, 1979.)

Nelson, D. R. y B. I. Halperin (1985), "Pentagonal and icosahedral order in rapidly cooled metals", *Science,* 229, p. 233.

Newton, I. (1687), *Principia,* Cambridge University Press. (Hay dos ediciones españolas, la primera preparada por A. Escotado, Editorial Nacional, 1980, y Tecnos, Madrid, 1988; y la segunda por E. Rada, Alianza Editorial, Madrid, 1988.)

Newton, I. (1730), *Opticks,* Dover Inc., 1952 (Hay edición española preparada por C. Solís, Alfaguara, Madrid, 1970.)

Oakley, D. A., (comp.) (1985), *Brain and mind,* Methuen, Londres y Nueva York.

Oakley, D. A. y L. C. Eames (1985), "The plurality of consciousness" en Oakley, D. A., (comp.) *Brain and mind,* Methuen, Londres y Nueva York.

O'Conneel, K. (1988), "Computer chess", *Chess,* 15.

O'Keefe, J. (1985), "¿Is consciousness the gateway to the hippocampal cognitive map? A especulative essay on the neutral basis of mind" en Oakley, D. A., (comp.) *Brain and mind,* Methuen, Londres y Nueva York.

Onoda, G. Y., P. J. Steinghardt, D. P. Di Vincenzo y J. E. S. Socolar, (1988), "Growing perfect quasicrystals", *Phys. Rev. Lett.,* 60, p. 2688.

Oppenheimer, J. R. y H. Snyder (1939), "On continued gravitational contraction", *Phys. Rev.* 56, pp. 455-459.

Pais, A. (1982), "*Subtle is the Lord..": the science and life of Albert Einstein,* Clarendon Press, Oxford. (Hay traducción española: *El Señor es sutil. La ciencia y la vida de Albert Einstein,* Editorial Ariel, Barcelona, 1984). [La carta de Einstein citada en la p. 280 figura en *Correspondencia Einstein-Born,* Siglo XXI, México, 1973, N. del T.]

París, J. y L. Harrington (1977), "A mathematical incompleteness in Peano arithmetic", en J. Barwise (comp.), *Handbook of mathematical logic,* North-Holland, Amsterdam.

Pearle, P. (1985), "Models for reduction", en C. J. Isham y R. Penrose (comps.), *Quantum concepts in space and time,* Oxford University Press.

____ (1989), "Combining stochastic dynamical state vector reduction with spontaneous localizations", *Phys. Rev. A,* 39, pp. 2277-2289.

Peitgen H.-O. y P. H. Richter (1986), *The beatuy of fractals,* Springer-Verlag, Berlín y Heidelberg.

____y D. Saupe (1988), *The science of fractals images,* Springer-Verlag, Berlin.

Penfield, W. y H. Hasper (1974), "Highest Level seizures", *Research Publications of the Association for Research in Nervous and Mental Diseases,* 26, Nueva York, pp. 252-271.

Penrose, R. (1965) "Gravitational collapse and space-time singularities", *Phys, Rev. Lett.,* 14, pp. 57-59.

____(1974), "The role of aesthetics in pure and applied mathematical research", *Bull. Inst. Math. Applications* 10, num. 7/8, pp. 266-271.

____(1979a), "Einstein's vision and the mathematics of the natural world", *The Sciences,* marzo, pp. 6-9.

____(1979b), "Singularities and time-asymmetry", en S. W. Hawking y W Israel (comps.), *General relativity: An Einstein centenary,* Cambridge, University Press.

____(1987a), "Newton, quantum theory and reality", en S. W. Hawking y W. Israel (comps.), *300 years of gravity,* Cambridge University Press.

____(1987b), "Quantum Physics and Conscious Thought", en B. J. Hiley y F D Peat (comps.), *Quantum implications, Essays in honour of David Bohm,* Routledge y Kegan Paul, Londres y Nueva York.

____(1989a), "Tilings and quasicrystals: a non-local growth problem?", en M. Jaric, *Aperiodicity and order 2,* Academic Press, Nueva York.

____(I989b), "Difficulties with inflationary cosmology", en *Proceeding of the 14th Texas Symposium on Relativistic Astropysics* (ed. por E. J. Fenyves), NY Acad. Sci., Nueva York.

____y W. Rindler (1984), *Spinors and space-time,* vol. I: *Two spinor calculus and relativistic fields,* Cambridge University Press.

____y W. Rindler (1986), *Spinors and space-time,* vol. II: *Spinor and twistor methods in space-time geometry,* Cambridge University Press. Pour-El, M.B. e I. Richards (1979), "A computable ordinary differential equation which possesses no computable solution", *Ann. Math. Logic,* 17, pp. 61-90.

____(1981), "The wave equation with computable initial data such that its unique solution is not computable", *Adv. in Math.,* 39, pp. 215-39.

____(1982), "Noncomputability in models of physical phenomena", *Int. J. Theor. Phys.,* 21, pp. 553-555.

____(1989), *Computability in analysis and physics,* Springer-Verlag, Nueva York.

Rae, A., (1986), *Quantum physics: illusion or reality?,* Cambridge University Press. (Hay traducción española: *Física cuántica: ¿ilusión o realidad?,* Alianza Editorial, Madrid, 1988.)

Resnikoff, H. L. y R. O. Jr. Wells (1973), *Mathematics and civilization,* Holt, Rinehart and Winston, Inc., Nueva York; reimpreso con adiciones, 1984, Dover Publications, Inc. Mineóla, Nueva York.

Rindler, W. (1977), *Essential relativity,* Springer-Verlag, Nueva York.

____(1982), *Introduction to special relativity,* Clarendon Press, Oxford.

Robinson, R. M. (1971), "Undecidability and nonperiodicity for tilings of the plane", *Invent. Math.,* 12, pp. 177-209.

Rouse Ball, W. W. (1992), "Calculating prodigies", en -*Mathematical recreations and essays.*

Rucker, R. (1984), *Infinity and the mind: the science and philosophy of the infinite,* Paladin Books, Granada Publishing Ltd., Londres; publicado originalmente por Birkhauser Inc., Boston, Mass., 1982.

Sachs, R. K. (1962), "Gravitational waves in general relativity VIII. Waves in asymptotically flat space-time", *Proc. Roy. Soc* Londres A 270, pp. 103-126.

Schank, R. C. y R. P. Abelson (1977), *Scripts, plants, goals and under standing,* Erlbaum, Hillsdale, New Jersey.

Schrödinger, E. (1935), "Die gegenwartige Situation in der Quantenmechanik, Naturwissenschaften", 23, pp. 807-812, 823-828 844-849 (Traducido al inglés por J. T. Trimmer en *Proc. Amer. Phil Soc* 124 1980, pp. 323-338; y enJ. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983

____(1967), *What is life and Mind and Matter,* Cambridge University Press. (Hay traducción española: *¿Qué es la vida?,* Editorial Tusquets, Barcelona, 1984, y Orbis, Barcelona, 1986; *Mente y materia,* Editorial Tusquets, Barcelona, 1985. También hay traducción catalana en Edicions 62, Barcelona, 1984.)

Searle, J. (1980), "Mind, brains and programs", en *The behavioral and brain sciences,* vol. 3, Cambridge University Press, reimpreso en D.R. Hofstadter y D.C. Dennett, *The mind's I,* Basic Books, Inc., Penguin Books Ltd., Harmondsworth, Middx, 1981.

____(1987), "Minds and brains without programs", en *Mindwaves* (ed. por C. Blakemore y S. Greenfield), Basil Blacwell, Oxford.

Shechtman, D., I. Blech, D. Gratias y J. W. Cahn (1984), "Metallic phase with long-range orientational order and no translational symmetry", *Phys. Rev. Lett.* 53, p. 1951.

Smith, S. B. (1983), *The great mental calculators,* Columbia University Press.

Smorynski, C. (1983), "'Big news from Archimedes to Friedman", *Notices Amer. Math. Soc.* 30, pp. 251-256.

Sperry, R. W. (1966), "Brain bisection and consciousness", en J. Eccles (comp.), *Brain and concious experience,* Springer, Nueva York.

Squires, E. (1985), *To acknowledge the wonder,* Adam Hilger Ltd., Bristol.

____(1986), *The mystery of the quantum world,* Adam Hilger Ltd., Bristol.

Tipler, F. J., C. J. S. Clarke y G. F. R. Ellis (1980), "Singularities and horizons —a review article", en A. Held (comp.), *General relativity and gravitation,* vol. II, Plenum Press, Nueva York, pp. 97-206.

Turing, A. M. (1937), "On computable numbers, with an aplication to the Entscheidungsproblem", *Proc. Lond. Math. Soc.* (ser. 2), 42, pp. 230-265; corrección en 43, pp.

____(1939), "Systems of logic based on ordinals", *Proc. Lond. Math. Soc.,* 45, pp. 161-228.

____(1950), "Computing machinery and intelligence", *Mind,* 59 num. 236; reimpreso en D. R. Hofstadter and D.C. Dennett, *The mind's I,* Basic Books, *Inc.,* Penguin Books, Ltd; Hardmonsworth, Middx, 1981. (Hay traducción española: "Maquinaria de cómputo e inteligencia", en Z W. Pylyshyn (comp.), *Perspectivas de la revolución de los computadores* Alianza Editorial, Madrid, 1975.)

Von Neumann, J. (1955), *Mathematical foundations of quantum mechanics,* Princeton University Press. (Hay traducción española: *Fundamentos matemáticos de la mecánica cuántica,* Instituto Jorge Juan csic Madrid, 1955.)

Waltz, D. L. (1982), "Artificial intelligence", *Scientific American,* 247 (4) pp. 101-122. (Traducción española en *Investigación y Ciencia,* diciembre de 1982, pp. 48-61.)

Ward, R. S. y R. O. Jr. Wells, (1990), *Twistor geometry and field theory,* Cambridge University Press.

Weinberg, S. (1977), *The first three minutes: A modern view for the origin of the universe,* André Deutsch, Londres. (Hay traducción española: *Los tres primeros minutos del Universo,* Alianza Editorial, Madrid, 1978),

Weiskrantz, L. (1987), "Neuropsychology and the nature of consciousness", en C. Blakemore y S. Greenfields (comps.), *Mindwaves* Black-well, Oxford.

Westfall, R. S. (1980), *Never at rest,* Cambridge University Press.

Wheeler, J. A. (1983), "Law without law", en J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, pp. 182-213.

____y R. P. Feynman (1945), "Interaction with the absorber as the mechanism of radiation", *Revs. Mod. Phys.,* 17, pp. 157-181.y W. H. Zurek (comps.) (1983), *Quantum theory and measurement,* Princeton University Press.

Whittaker, E. T. (1910), *The history of the theories of aether and electricity,* Longman, Londres. [Reeditado junto con la segunda parte, en Dover Publishing, Inc., 1989 (N. del T.)]

Wigner, E. P. (1960), "The unreasonable effectiveness of mathematics", *Commun. Pure Appl. Math.* 13, pp. 1-14.

____(1961), "Remarks on the mind body question", en I. J. Good, *The scientist speculates,* Heinemann, Londres. (Reimpreso en E. Wigner *Symmetries and reflections,* Indiana University Press, Bloomington, 1967; y J. A. Wheeler y W. H. Zurek (comps.), *Quantum theory and measurement,* Princeton University Press, 1983.)

Will, C. M. (1987), "Experimental gravitation from Newton's Principia to Einstein's general relativity", en S. W. Hawking y W. Israel (comps.) *300 years of gravitation,* Cambridge University Press.

Wilson D H., A. G. Reeves, M. S. Gazzaniga y C. Culver (1977), "Cerebral commissurotomy for the control of intractable seizures", *Neurology* 27 pp 708-715

Winograd, T. (1972), "Understanding natural languaje", *Cognitive Psychology* 3, pp. 1-191.

Wooters W K. y W. H. Zurek (1982), "A single quantum cannot be cloned", *Nature,* 299, pp. 802-803.

ÍNDICE GENERAL

Este libro se terminó de imprimir y encuadernar en el mes de diciembre de 19% en Impresora y Encuadernadora Progreso, S.A. de C.V. (IEPSA), Calz. de San Lorenzo, 244; 09830 México, D.F. Se tiraron 2 000 ejemplares

Revisión de la traducción: *Marcelino Perelló* e *Ignacio Aguilar Marcué.*

Tipografía y formación: *José Luis Acosta,* de επιστημη

Preparación de índice analítico: *Marcela Pimentel.*

Preparación de originales mecánicos: *Araceli Vázquez.*

Fotomecánica: *Formación Gráfica.*

Cuidó la edición *Axel Retif.*

Esta obra es una coedición del CONSEJO NACIONAL DE CIENCIA Y TECNOLOGÍA y el FONDO DE CULTURA ECONÓMICA, coordinada por *María del Carmen Farías.*